STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 107598

TO: Celine Qian **Location: CM1/11C01&11E12** **Art Unit: 1636** **Wednesday, November 12, 2003** **Case Serial Number: 09/939209**	**From: Edward Hart** **Location: Biotech-Chem Library** **CM1-6B02** **Phone: 305-9203** **edward.hart@uspto.gov**

Search Notes

Examiner Qian,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 01:54:16 ; Search time 3916.96 Seconds
(without alignments)
10444.222 Million cell updates/sec

Title: US-09-939-209A-3_COPY_1_1000
Perfect score: 1000
Sequence: 1 agttcaagaccagcctgagc.....gtatgtggcagaggcatggt 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1000	100.0	20300	6	AX451337	AX451337 Sequence
2	1000	100.0	165329	9	AL583850	AL583850 Human DNA
c 3	617.4	61.7	191699	2	AC031977	AC031977 Homo sapi
4	411.2	41.1	203216	2	AC097328	AC097328 Pan trogl
c 5	405.2	40.5	181946	2	AC097330	AC097330 Pan trogl
6	402.2	40.2	109081	9	AL358074	AL358074 Human DNA
7	396.8	39.7	164844	9	AC009058	AC009058 Homo sapi
8	395.2	39.5	130349	9	AC011593	AC011593 Homo sapi
c 9	394.8	39.5	196413	2	AC097269	AC097269 Pan trogl
10	390.8	39.1	189662	9	AC015468	AC015468 Homo sapi
c 11	387.4	38.7	125685	9	AC073310	AC073310 Homo sapi
12	385.8	38.6	139118	9	AL133477	AL133477 Human DNA
13	382.6	38.3	216021	9	HUAC004787	AC004787 Homo sapi
14	379.8	38.0	92564	9	AY007685	AY007685 Homo sapi
c 15	379.4	37.9	161252	2	AC114955	AC114955 Homo sapi
c 16	378	37.8	202305	9	AC114291	AC114291 Homo sapi
c 17	373.6	37.4	192139	9	AC021955	AC021955 Homo sapi
c 18	372	37.2	99596	9	AC106873	AC106873 Homo sapi
c 19	371.4	37.1	144383	2	BX324168	BX324168 Homo sapi
20	370.4	37.0	252512	2	AC097326	AC097326 Pan trogl
21	368.2	36.8	170646	2	AC123545	AC123545 Pan trogl
22	367	36.7	112762	9	HS323B6	Z83841 Human DNA s
c 23	365.4	36.5	163976	9	AC108483	AC108483 Homo sapi
24	365.4	36.5	163998	9	AC022166	AC022166 Homo sapi
c 25	364.2	36.4	153201	2	AC092863	AC092863 Homo sapi
c 26	363.6	36.4	151549	9	AC092694	AC092694 Homo sapi
27	357.4	35.7	217985	9	AC021193	AC021193 Homo sapi
c 28	356.6	35.7	121600	9	HS141H5	AL049176 Human DNA
c 29	355.2	35.5	149969	2	AC024180	AC024180 Homo sapi
c 30	355.2	35.5	162996	9	AC006441	AC006441 Homo sapi
c 31	354.8	35.5	182430	9	AL590381	AL590381 Human DNA
c 32	353.6	35.4	164994	2	AC141288	AC141288 Homo sapi
33	352.4	35.2	178254	9	AL731541	AL731541 Human DNA
c 34	349.6	35.0	50612	9	AL513530	AL513530 Human DNA
35	349.6	35.0	81696	9	AF438327_3	Continuation (4 of
36	349.6	35.0	110000	9	AF438327_2	Continuation (3 of
c 37	349.2	34.9	171427	9	AC021590	AC021590 Homo sapi
c 38	348.8	34.9	236551	2	AC094041	AC094041 Rattus no
39	348.6	34.9	134084	9	AC004966	AC004966 Homo sapi
c 40	348.6	34.9	159817	2	BX322798	BX322798 Danio rer
c 41	348.4	34.8	247615	2	AC098061	AC098061 Rattus no
42	347.8	34.8	220821	2	AC133068	AC133068 Danio rer
c 43	347.4	34.7	203216	2	AC097328	AC097328 Pan trogl
44	347.2	34.7	174447	2	AC142552	AC142552 Danio rer
45	347	34.7	137325	2	AC142088	AC142088 Danio rer

ALIGNMENTS

RESULT 1
AX451337
LOCUS AX451337 20300 bp DNA linear PAT 03-JUL-2002
DEFINITION Sequence 3 from Patent WO0216653.
ACCESSION AX451337
VERSION AX451337.1 GI:21698388
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Levitt,P.R., Mirnics,K., Kodavali,V.C. and Nimgaonkar,V.L.
TITLE Methods and systems for facilitating the diagnosis and treatment of schizophrenia
JOURNAL Patent: WO 0216653-A 3 28-FEB-2002;

```
          University of Pittsburgh (US)
FEATURES             Location/Qualifiers
     source          1. .20300
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="A genomic sequence containing RGS4 nucleic acid
                     sequence and sequ ences upstream and downstream to the
                     RGS4 nucleic acid sequence"
BASE COUNT     6157 a   4102 c   3775 g   6266 t
ORIGIN

  Query Match           100.0%;  Score 1000;  DB 6;  Length 20300;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-185;
  Matches 1000;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATATACA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480

Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540

Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780

Qy        781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840

Qy        841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900

Qy        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960

Qy        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGT 1000
              |||||||||||||||||||||||||||||||||||||||
Db        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGT 1000


RESULT 2
AL583850
LOCUS       AL583850              165329 bp    DNA     linear   PRI 15-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-430G6 on chromosome 1, complete
            sequence.
ACCESSION   AL583850
VERSION     AL583850.5  GI:16973044
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165329)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 16, 2001 this sequence version replaced gi:15020514.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-430G6 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-430G6 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP11-430G6 is at 165329 in this
            sequence. The true right end of clone RP11-331H2 is at 2000 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .165329
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-430G6"
                     /clone_lib="RPCI-11.2"
BASE COUNT    51144 a  30897 c  31439 g  51849 t
ORIGIN
```

```
  Query Match             100.0%;  Score 1000;  DB 9;  Length 165329;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-185;
  Matches 1000;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71683 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 71742

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71743 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 71802

Qy        121 CACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71803 CACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 71862

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71863 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 71922

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71923 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 71982

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71983 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 72042

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72043 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 72102

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72103 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 72162

Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72163 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 72222

Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72223 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 72282

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72283 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 72342

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72343 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 72402

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72403 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 72462

Qy        781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72463 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 72522

Qy        841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72523 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 72582

Qy        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72583 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 72642

Qy        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGT 1000
              |||||||||||||||||||||||||||||||||||||||
Db      72643 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGT 72682
```

```
RESULT 3
AC031977/c
LOCUS       AC031977              191699 bp    DNA     linear   HTG 12-APR-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-288O18, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC031977
VERSION     AC031977.7  GI:13194952
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
            Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
            Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
            Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Mar 4, 2001 this sequence version replaced gi:9665085.
            -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 880
                 Center clone name: RP11-288O18
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513; 100% of reads
                 Sequencing Vector: plasmid; plasmid_accession; 0% of reads
                 Chemistry: Dye-primer; 1% of reads
                 Chemistry: Dye-terminator Big Dye; 99% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 190680 bases at least Q40
                 Consensus quality: 191287 bases at least Q30
                 Consensus quality: 191336 bases at least Q20
                 Insert size: 195548; agarose-fp
                 Insert size: 191499; sum-of-contigs
                 Quality coverage:  7.9x in Q20 bases; agarose-fp
                 Quality coverage:  8.1x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    12646: contig of 12646 bp in length
            *    12647    12746: gap of unknown length
            *    12747    94961: contig of 82215 bp in length
            *    94962    95061: gap of unknown length
            *    95062   191699: contig of 96638 bp in length.
FEATURES            Location/Qualifiers
     source         1. .191699
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="1"
                    /clone="RP11-288O18"
                    /clone_lib="RPCI human BAC library 11"
```

```
     misc_feature    1. .12646
                     /note="assembly_name:Contig5"
     misc_feature    12747. .94961
                     /note="assembly_name:Contig6
                     clone_end:T7"
     misc_feature    95062. .191699
                     /note="assembly_name:Contig7
                     clone_end:SP6"
BASE COUNT    57654 a  36385 c  36166 g  61293 t    201 others
ORIGIN

  Query Match             61.7%;  Score 617.4;  DB 2;  Length 191699;
  Best Local Similarity   82.9%;  Pred. No. 3.2e-111;
  Matches  677;  Conservative    0;  Mismatches  126;  Indels   14;  Gaps    3;

Qy        194 GCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATAC 253
              |||
Db      12749 GCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 12690

Qy        254 ACACACACACACATATATATATACACATATATATACACATATATACATATATACACATAT 313
                                                  |||| ||||| | | |
Db      12689 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTACACATATATATACAC 12630

Qy        314 ATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATAT 373
              |||   ||| | ||| | | |||||| |||||| |||||||||||||| || ||| ||||
Db      12629 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACATATAT 12570

Qy        374 ATACACATATATACACACATATATACACATATATACACACATATATACACATATATACAC 433
              | |||||||||||||||||||||||||||||||    |||||||||||||||||||||||||
Db      12569 ACACACATATATACACACATATATACACATAT----ACACATATATACACATATATACAC 12514

Qy        434 ATATATACACATATACACATATATACACATATATACATATATACACATATATATAATATA 493
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      12513 ATATATACACATATATACATATATACACATATATACATATATACACATATATATAATATA 12454

Qy        494 CACACATATATATACACATATATACACACATATATACACATATATACACATAT------A 547
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||      |
Db      12453 CACACATATATATACACATATATACACACATATATACACATATATACACATATATACACA 12394

Qy        548 TATACACATATATACACATATATACATATATACACATATATATACATATATACACATATA 607
              |||| |||||||||||||||||| | ||||| ||| |||||||||||||||||||||||||
Db      12393 TATATACATATATACACATATATATACATATATACACATATATACATATATACACATATA 12334

Qy        608 TACATATATACACATATATACATATATACACACATATATACACATACATATACACACACA 667
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12333 TACATATATACACATATATACATATATACACACATATATACACATACATATACACACACA 12274

Qy        668 TAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAG 727
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db      12273 TAGATATACATATATATACACATATATATGTATATATATGTATATATATATGCTCCAG 12214

Qy        728 AGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGC 787
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12213 AGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGC 12154

Qy        788 AGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATG 847
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12153 AGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATG 12094

Qy        848 TGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACAC----TATATATATATATAG 903
              |||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db      12093 TGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATATATAG 12034

Qy        904 TGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTA 963
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12033 TGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTA 11974

Qy        964 AGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGT 1000
              |||||||||||||||||||||||||||||||||||||
Db      11973 AGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGT 11937

RESULT 4
```

```
AC097328
LOCUS       AC097328               203216 bp    DNA     linear   HTG 20-OCT-2001
DEFINITION  Pan troglodytes clone RP43-174K14, WORKING DRAFT SEQUENCE, 21
            unordered pieces.
ACCESSION   AC097328
VERSION     AC097328.1  GI:16118151
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 203216)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 203216)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     ---------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
              Center project name: ZUAM
              Center clone name: RP43-174K14
            ---------- Summary Statistics
              Sequencing vector: Plasmid; M77789
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 196812 bases at least Q40
              Consensus quality: 208422 bases at least Q30
```

```
    Consensus quality: 218080 bases at least Q20
    Estimated insert size: 228347; sum-of-contigs estimation
    Quality coverage: 0x in Q20 bases; agarose-fp estimation
    Quality coverage: 7.1x in Q20 bases; sum-of-contigs estimation
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1    30770: contig of 30770 bp in length
*    30771    30870: gap of unknown length
*    30871    52356: contig of 21486 bp in length
*    52357    52456: gap of unknown length
*    52457    72697: contig of 20241 bp in length
*    72698    72797: gap of unknown length
*    72798    90396: contig of 17599 bp in length
*    90397    90496: gap of unknown length
*    90497    96134: contig of 5638 bp in length
*    96135    96234: gap of unknown length
*    96235   114433: contig of 18199 bp in length
*   114434   114533: gap of unknown length
*   114534   125184: contig of 10651 bp in length
*   125185   125284: gap of unknown length
*   125285   135194: contig of 9910 bp in length
*   135195   135294: gap of unknown length
*   135295   140273: contig of 4979 bp in length
*   140274   140373: gap of unknown length
*   140374   148490: contig of 8117 bp in length
*   148491   148590: gap of unknown length
*   148591   153435: contig of 4845 bp in length
*   153436   153535: gap of unknown length
*   153536   162229: contig of 8694 bp in length
*   162230   162329: gap of unknown length
*   162330   167843: contig of 5514 bp in length
*   167844   167943: gap of unknown length
*   167944   175297: contig of 7354 bp in length
*   175298   175397: gap of unknown length
*   175398   179409: contig of 4012 bp in length
*   179410   179509: gap of unknown length
*   179510   185931: contig of 6422 bp in length
*   185932   186031: gap of unknown length
*   186032   190690: contig of 4659 bp in length
*   190691   190790: gap of unknown length
*   190791   194808: contig of 4018 bp in length
*   194809   194908: gap of unknown length
*   194909   198788: contig of 3880 bp in length
*   198789   198888: gap of unknown length
*   198889   201055: contig of 2167 bp in length
*   201056   201155: gap of unknown length
*   201156   203216: contig of 2061 bp in length.
FEATURES             Location/Qualifiers
     source          1. .203216
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-174K14"
BASE COUNT    59570 a  41381 c  40477 g  59769 t   2019 others
ORIGIN

  Query Match             41.1%;  Score 411.2;  DB 2;  Length 203216;
  Best Local Similarity   76.7%;  Pred. No. 4.3e-71;
  Matches  554;  Conservative    0;  Mismatches  163;  Indels    5;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCA-TCTCTACTAAAAATACAAAATT 59
              ||||||| |||||||||| |||||||||||||||||  |||||||||||| ||||||||
Db     114617 AGTTCAAAACCAGCCTGACCAACATGGTGAAACCCCCGTCTCTACTAAAAACACAAAATT 114676

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   |||||||||| | || ||||||| |||||||||| ||| ||  ||||||||||
Db     114677 AGCTGGGCATGGTGGTGCATGCCTGTAGTCCCAGCTACTCAGGAAGCTAAGGCAGGAGAA 114736

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              |||||||||||  | || || ||||||||| | || |  |||||||||||| |||||||||
Db     114737 TCACTTGAACC--CAGGAGGCGGAGGTTGCAGTGAACTGAGATCATGCCACTGCACTCCA 114794

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||  | ||||| |||         |  |    | | |   |       | | |||||||||
Db     114795 GCCTGAGTAACAAAAGCATCTCAAAAAATATATATATATGTATATATGTGTGTATATATA 114854

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              | | |||||||||         | | | |  |||||||||    | | |||| |  | ||||
Db     114855 TGTGTATATATATGTGTGTATATATATGTGTATATATATGTGTGTGTATATATGTGTATA 114914

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              |||||||  | |||||||  | |||||||||  | |||  |||| ||  | |||||||||  | |
Db     114915 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 114974

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              |||  ||||  | |||||||  | |||||||  | | |||||||  | |||||||  | |  |||||
Db     114975 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115034

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              || | |||||||  | |||||||  | |||||  | |||||||  | |||||||  |||||||  |
Db     115035 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115094

Qy        479 CATATATATAATATACACACATATATATACACATATATACACACATATATACACATATAT 538
              ||||||||  |||||  | |  |||||||||  | |||||||  | |  |||||||  | ||||||
Db     115095 TATATATAT-ATATATATATATATATATATATATATATATATATATATATATATATAT 115153

Qy        539 ACACATATATATACACATATATACACATATATACACATATATACACATATATATACATATAT 598
              | | |||||||||  | |||||||  | |||||||  |||||||  | |||||||||  ||||||
Db     115154 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115213

Qy        599 ACACATATATACATATATACACATATATACATATATACACACATATATACACATACATAT 658
              | | ||||||| |||||||  | |||||||  |||||||  | | |||||||  | |||  ||||
Db     115214 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115273

Qy        659 ACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATAT 718
              | | |   ||| |||||  |||||||||  | |||||||||  |||||||||  |||||||||
Db     115274 ATATATCTATATATATATATATATATATATATATATATATATATATATATATATATAT 115333

Qy        719 AT 720
              ||
Db     115334 AT 115335


RESULT 5
AC097330/c
LOCUS       AC097330               181946 bp    DNA     linear   HTG 20-OCT-2001
DEFINITION  Pan troglodytes clone RP43-54O15, WORKING DRAFT SEQUENCE, 16
            unordered pieces.
ACCESSION   AC097330
VERSION     AC097330.1  GI:16118153
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 181946)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
```

```
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181946)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: ZUAP
                Center clone name: RP43-54O15
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 169928 bases at least Q40
                Consensus quality: 176100 bases at least Q30
                Consensus quality: 180968 bases at least Q20
                Estimated insert size: 183838; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 6.9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    17715: contig of 17715 bp in length
            *    17716    17815: gap of unknown length
            *    17816    44117: contig of 26302 bp in length
            *    44118    44217: gap of unknown length
            *    44218    67185: contig of 22968 bp in length
            *    67186    67285: gap of unknown length
            *    67286    84740: contig of 17455 bp in length
            *    84741    84840: gap of unknown length
            *    84841   100199: contig of 15359 bp in length
            *   100200   100299: gap of unknown length
            *   100300   112190: contig of 11891 bp in length
            *   112191   112290: gap of unknown length
            *   112291   124548: contig of 12258 bp in length
            *   124549   124648: gap of unknown length
            *   124649   134134: contig of 9486 bp in length
            *   134135   134234: gap of unknown length
            *   134235   143821: contig of 9587 bp in length
            *   143822   143921: gap of unknown length
            *   143922   152987: contig of 9066 bp in length
            *   152988   153087: gap of unknown length
            *   153088   159063: contig of 5976 bp in length
            *   159064   159163: gap of unknown length
            *   159164   166547: contig of 7384 bp in length
            *   166548   166647: gap of unknown length
            *   166648   171663: contig of 5016 bp in length
            *   171664   171763: gap of unknown length
            *   171764   177471: contig of 5708 bp in length
            *   177472   177571: gap of unknown length
            *   177572   179822: contig of 2251 bp in length
            *   179823   179922: gap of unknown length
            *   179923   181946: contig of 2024 bp in length.
FEATURES                Location/Qualifiers
     source             1. .181946
                        /organism="Pan troglodytes"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9598"
                        /clone="RP43-54O15"
BASE COUNT    49474 a  40705 c  41770 g  48491 t   1506 others
ORIGIN

  Query Match            40.5%;  Score 405.2;  DB 2;  Length 181946;
  Best Local Similarity  68.3%;  Pred. No. 6.4e-70;
  Matches  634;  Conservative    0;  Mismatches  288;  Indels    6;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||  ||||||  ||||   | |  |||||||||||||  |||||||||  ||||||  ||||||
Db     155131 AGTTCGAGACCAACCTGGACTATATGGTGAAACCCCGTCTCTACTAGAAATACAAAAATT 155072

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   |||||||||  ||||||||||||||||||  ||||||  ||||||   ||||||||||
Db     155071 AGCTGGGCATGGTGGCACACGCCTGTAATCCCAACTACTTGGGAGGCTAAGGCAGGAGAA 155012

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              ||||||||||    | ||  ||  ||||||||  | |||  |  |||||    |||  |||||||||
Db     155011 TCACTTGAAC--TCGGGAGGCAGAGGTTGCAGTGAGCCCAGATCGCACCACTGCACTCCA 154954

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||  ||||  |       |    ||  |||     |  ||||  ||||  |     |       |  | |
Db     154953 GCCTAGGCGAGTGAGACTCCATCTCAAAAAAAAAAAAGAAAAGAAAATTGAGGGCTGGACA 154894

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              |      ||  |      |  |  ||    |  |  |||||  |  |||||||||  |  |||||||
Db     154893 CAGTGGCTTACACCTGTAATCCCAGCACTTTGGATATATATATATATATATATATATATA 154834

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              |||||||  |  |||||||  |  |||||||||  |  |||  ||||  ||  |  |||||||||  |  |
Db     154833 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154774

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              |||  ||||  |  |||||||  |  |||||||  |  |  |||||||  |  |||||||  |  |  |||||
Db     154773 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154714

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              ||  |  |||||||  |  |||||||  |  |||||  |  |||||||  |  |||||||  |||||||  |
Db     154713 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154654

Qy        479 CATATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||||||  |||||  |  |  |||||||||  |||||||  |  |  |||||||  |  |||||
Db     154653 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154594
```

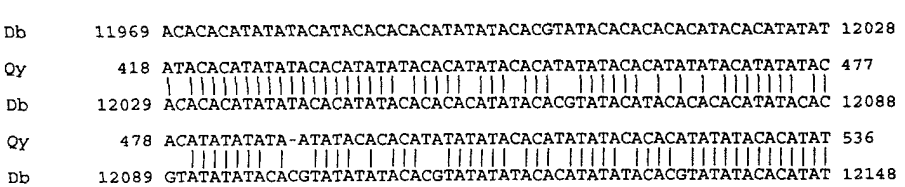
```
Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              || | ||||||||| | ||||||| | ||||||| ||||||| | ||||||||| |||||
Db     154593 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154534

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| |||
Db     154533 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154474

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | | ||| ||||| ||||||||| | |||||||||  ||||||||| |||||||||
Db     154473 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154414

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              |||          | |||          | |   ||    |    ||| |   | | | |
Db     154413 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154354

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
                | | |    | |||    | |   | || |      |      | |||       | |
Db     154353 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154294

Qy        838 TACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATA 896
              || ||| || | ||   | |  ||| | || |  ||  ||   |  ||    ||||||||
Db     154293 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154234

Qy        897 TATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||||||   | |    | || |  || |
Db     154233 TATATATATATATATATATATATATA 154206
```

RESULT 6
AL358074
LOCUS AL358074 109081 bp DNA linear PRI 21-NOV-2001
DEFINITION Human DNA sequence from clone RP11-423C15 on chromosome 9, complete sequence.
ACCESSION AL358074
VERSION AL358074.22 GI:17048358
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 109081)
AUTHORS Hammond,S.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Nov 22, 2001 this sequence version replaced gi:17043779.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr9
RP11-423C15 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-423C15 It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true right end of clone RP11-423C15 is at 109081 in this sequence. The true left end of clone RP11-770B5 is at 78953 in this sequence. The true right end of clone RP11-12A16 is at 2000 in this sequence.

```
FEATURES             Location/Qualifiers
     source          1. .109081
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-423C15"
                     /clone_lib="RPCI-11.2"
     misc_feature    complement(65197. .65223)
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    65533. .65757
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    98714. .98750
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    98834. .98842
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    99111. .99161
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly"
     misc_feature    99163. .99191
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly"
BASE COUNT    31179 a  22615 c  23183 g  32104 t
ORIGIN
```

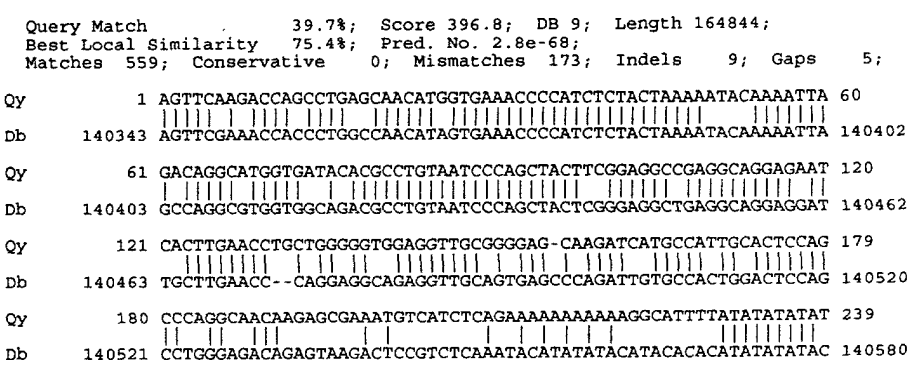
Query Match 40.2%; Score 402.2; DB 9; Length 109081;
Best Local Similarity 78.2%; Pred. No. 2.5e-69;
Matches 564; Conservative 0; Mismatches 138; Indels 19; Gaps 6;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              ||||  |||||||||||  ||||||||||||||| ||||||||||||||||  ||||
Db      11620 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAC 11679

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              ||| || || |||||  || | ||||||||||||||||||  ||  || ||||||||||||
Db      11680 TAGCCAAGCGTGGTGGCACGCACCTGTAATCCCAGCTACTCGGGCAGCTGAGGCAGGAGA 11739

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| |||||||  |  | ||  |||||||| | ||| ||||||||||||||||||||||||
Db      11740 ATCGCTTGAACC--CAAGAGGCAGAGGTTGCAGTGAGCCAAGATCATGCCATTGCACTCC 11797

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||| || | ||||||||||||||  |  |||| ||||||||||        ||||||||
Db      11798 AGCTCAAGTAACAAGAGCGAAATTCCGCCTCAAAAAAAAAAAA---------ATATATAT 11848

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              |||||||| | | |||||||| |||  |||||||| ||||| ||||| ||| |||||
Db      11849 ATATATATACACACACACACACATACATGTATATATACACACACATATACACATATATAC 11908

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||| ||||| ||| ||| ||||| ||||||||||||| | |   ||| ||||| ||| ||
Db      11909 ACACATATATACACATACACACACATATATACACATATACACACACATATATACATATAC 11968

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              | |   ||| | | ||| ||||| ||||| |||| |||| ||||| | ||||||| ||||
```



RESULT 7
AC009058
LOCUS AC009058 164844 bp DNA linear PRI 13-MAR-2003
DEFINITION Homo sapiens chromosome 16 clone RP11-292B23, complete sequence.
ACCESSION AC009058
VERSION AC009058.10 GI:28933531
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 164844)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 164844)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 164844)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 164844)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 5 (bases 1 to 164844)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Mar 13, 2003 this sequence version replaced gi:17432456.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos National Laboratory
www-shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
FEATURES Location/Qualifiers
source 1. .164844
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-292B23"
BASE COUNT 47164 a 37566 c 36346 g 43768 t
ORIGIN

Query Match 39.7%; Score 396.8; DB 9; Length 164844;
Best Local Similarity 75.4%; Pred. No. 2.8e-68;
Matches 559; Conservative 0; Mismatches 173; Indels 9; Gaps 5;



RESULT 8
AC011593
LOCUS AC011593 130349 bp DNA linear PRI 07-NOV-2001
DEFINITION Homo sapiens BAC clone RP11-437E9 from 2, complete sequence.
ACCESSION AC011593
VERSION AC011593.8 GI:14190772
KEYWORDS HTG.

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 130349)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 130349)
  AUTHORS   Pearman,C. and Cotton,M.
  TITLE     The sequence of Homo sapiens BAC clone RP11-437E9
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 130349)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 130349)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 130349)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On May 23, 2001 this sequence version replaced gi:13518261.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0437E09
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-289J14, 200 bp overlap; the
            clone sequenced to the right is AC061984.  Actual start of this
            clone is at base position 150577 of RP11-289J14; actual end is at
            base position 130349 of RP11-437E9.
FEATURES             Location/Qualifiers
     source          1. .130349
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-437E9"
                     /clone_lib="RPCI-11"
     repeat_region   1813. .2297
                     /rpt_family="MaLR"
     repeat_region   2618. .3136
                     /rpt_family="MaLR"
     repeat_region   3200. .3370
                     /rpt_family="MaLR"
     repeat_region   3522. .3856
                     /rpt_family="Alu"
     repeat_region   3866. .4290
                     /rpt_family="MaLR"
     repeat_region   4325. .4805
                     /rpt_family="ERV1"
     repeat_region   4910. .5014
                     /rpt_family="MER1_type"
     repeat_region   5258. .5600
                     /rpt_family="L1"
     repeat_region   5601. .5772
                     /rpt_family="Alu"
     repeat_region   5773. .6644
                     /rpt_family="L1"
     repeat_region   6773. .7356
                     /rpt_family="L1"
     repeat_region   7389. .7516
                     /rpt_family="L1"
     repeat_region   7529. .8188
                     /rpt_family="L1"
     repeat_region   8189. .8490
                     /rpt_family="Alu"
     repeat_region   8491. .8590
                     /rpt_family="L1"
     repeat_region   8605. .8943
                     /rpt_family="L1"
     repeat_region   9017. .9042
                     /rpt_family="(CA)n"
     repeat_region   9062. .9104
                     /rpt_family="MER53"
     repeat_region   9105. .9394
                     /rpt_family="Alu"
     repeat_region   9395. .9508
                     /rpt_family="MER53"
     repeat_region   9786. .9966
                     /rpt_family="C-rich"
     repeat_region   9946. .10036
                     /rpt_family="(TCC)n"
     repeat_region   10080. .10374
                     /rpt_family="Alu"
     repeat_region   12491. .12521
                     /rpt_family="(CA)n"
     repeat_region   13138. .13321
                     /rpt_family="MIR"
     repeat_region   13414. .13442
                     /rpt_family="(TAA)n"
     repeat_region   13615. .13925
                     /rpt_family="Alu"
     repeat_region   16183. .16266
                     /rpt_family="MIR"
     repeat_region   16629. .16997
                     /rpt_family="MaLR"
     repeat_region   17643. .17821
                     /rpt_family="Tc2"
     repeat_region   18157. .18200
                     /rpt_family="(TG)n"
```

```
  repeat_region   19591. .19619
                  /rpt_family="AT_rich"
  repeat_region   21204. .21375
                  /rpt_family="MaLR"
  repeat_region   21376. .21940
                  /rpt_family="ERVL"
  repeat_region   21962. .22901
                  /rpt_family="L1"
  repeat_region   22909. .23443
                  /rpt_family="L1"
  repeat_region   23492. .23987
                  /rpt_family="L1"
  repeat_region   23988. .24016
                  /rpt_family="(TTTTG)n"
  repeat_region   24017. .24634
                  /rpt_family="L1"
  repeat_region   24636. .25726
                  /rpt_family="L1"
  repeat_region   25727. .26040
                  /rpt_family="Alu"
  repeat_region   26041. .27537
                  /rpt_family="L1"
  repeat_region   27607. .28123
                  /rpt_family="ERVL"
  repeat_region   28282. .28808
                  /rpt_family="ERVL"
  repeat_region   28809. .29192
                  /rpt_family="MaLR"
  repeat_region   29193. .29577
                  /rpt_family="ERVL"
  repeat_region   29628. .29895
                  /rpt_family="Alu"
  repeat_region   29960. .30264
                  /rpt_family="ERVL"
  repeat_region   30265. .30573
                  /rpt_family="Alu"
  repeat_region   30574. .30753
                  /rpt_family="ERVL"
  repeat_region   30774. .32056
                  /rpt_family="ERVL"
  repeat_region   32100. .32242
                  /rpt_family="ERVL"
  repeat_region   32477. .32750
                  /rpt_family="MaLR"
  repeat_region   32751. .32847
                  /rpt_family="L1"
  repeat_region   32848. .33164
                  /rpt_family="Alu"
  repeat_region   33165. .33302

 Query Match             39.5%;  Score 395.2;  DB 9;  Length 130349;
 Best Local Similarity   67.8%;  Pred. No. 5.7e-68;
 Matches  645;  Conservative    0;  Mismatches  293;  Indels   14;  Gaps    6;

Qy          4 TCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAATTAG 61
              ||||||||| ||||   |||||||||||| |||||||||||||||||||||  ||||||||
Db      39934 TCAAGACCATCCTGGCTAACATGGTGAATCCCCATCTCTACTAAAAATACAAAAAATTAG 39993

Qy         62 ACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAATC 121
              ||||| |||||       ||||||| |||||||||||  ||||||  ||||||||||||
Db      39994 CCAGGCGTGGTGGCGGGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATG 40053

Qy        122 ACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCC 181
              | ||||||  | || ||||||| ||||   |||| |||||   ||||||||||||||||
Db      40054 GCGTGAACC--CGGGAGGTGGAGCTTGCAATGAGCCAGATCGCACCATTGCACTCCAGCC 40111

Qy        182 CAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTT---TATATATATA 238
              | ||||  |||| | |   ||||| | ||| | | | |   || |   || |||||||
Db      40112 TGGACAAC-GGAGCAAGACTCCATCTAAAAAATATATATAAATATATACATAAATATATA 40170

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              |||||| ||||| ||| | | | | ||||| ||||| ||||| ||||| |   |||| |
Db      40171 TATATAAATATAGACATAAATATATATATATAAATATAGACATAAATATATATGCATAAA 40230

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              ||||||| |||| ||| | ||| | ||| | ||| |||| |||| | ||||||| | |
Db      40231 TATATATATGCATAAATATATATAAAAATATATATAAATATATACATAAATATATATAAA 40290

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||| || | | ||||||| | ||||||| | | ||||||| | |||||||||||| |||||
Db      40291 TATATACAAAAATATATATAAATATATAAAAAAATATATAAATATATATACACATATATA 40350

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              | | ||||| | ||||||||| |||||| ||||| ||| | || |||| |||||||| |
Db      40351 AATATATATACATACATATATAAACATATATACATAAATATATATGTATAAATATATATA 40410

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              || ||| |||    |||  | | ||||| ||| | ||||||  | | |||||||| |||||
Db      40411 CACATAAATATATGTATGAATATATATACATAAATATATATGTATAAATATATATACATA 40470

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              ||| | | | ||||| ||||| ||| | ||| ||| ||||| | | ||||||||| |||
Db      40471 AATATATAAAGATATATACATAAATATATATAAATATATATACATAAATATATATAAATA 40530

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              |||| | ||| ||| ||| ||| | ||||| | ||||| | | | ||||| | | ||| |
Db      40531 TATATAAATAGATATATAAATATATATATAAATATATAAATATATATATAAATATATAAA 40590

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              |||| | | | | | ||||| |||||||| | |||||||  |  ||||||||  ||||||
Db      40591 TATATAAAAATAGATATATAAATATATATATAAATATATAAATATATATATAAATATATA 40650

Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775
              || ||    |   |  |||  |       |      || | |    ||| |   | | |
Db      40651 TAAATATATAAATATATATATAAATATATATAAATATATAAATATATATAAATATATAAA 40710

Qy        776 GAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGA 835
              | |  | |   | |||   | |   | || |   |      |||||       |
Db      40711 TATATATATAAATATATATAAATATATAAATATATATAAATATATATAAATATATAAATA 40770

Qy        836 AATACATAAATGTGTAATAAAACCTATCTGTA---AGGTAGTTAAGAAGGTAACACTATA 892
              ||| |||||| | ||  || |  ||  | ||   |  |  ||  ||  ||  | ||||
Db      40771 TATATATAAATATATATAAATATATAAATATATATATAAATATATAAATATATAAATATA 40830

Qy        893 TATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
              |||||| |||   |||    | ||||  || |  |       ||  || ||| |
Db      40831 TATATAAATATATAAATATATATAAATATATATAAATATATAAATATATATA 40882

RESULT 9
AC097269/c
LOCUS       AC097269               196413 bp    DNA     linear   HTG 20-OCT-2001
DEFINITION  Pan troglodytes clone RP43-11J15, WORKING DRAFT SEQUENCE, 18
            unordered pieces.
ACCESSION   AC097269
VERSION     AC097269.1  GI:16117532
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 196413)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
```

```
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196413)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
              Center project name: ZUAW
              Center clone name: RP43-11J15
            -------------- Summary Statistics
              Sequencing vector: Plasmid; M77789
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 190674 bases at least Q40
              Consensus quality: 194102 bases at least Q30
              Consensus quality: 196834 bases at least Q20
              Estimated insert size: 194298; sum-of-contigs estimation
              Quality coverage: 0x in Q20 bases; agarose-fp estimation
              Quality coverage: 8.1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    17120: contig of 17120 bp in length
            *    17121    17220: gap of unknown length
            *    17221    36306: contig of 19086 bp in length
            *    36307    36406: gap of unknown length
            *    36407    58958: contig of 22552 bp in length
            *    58959    59058: gap of unknown length
            *    59059    82373: contig of 23315 bp in length
            *    82374    82473: gap of unknown length
            *    82474    96867: contig of 14394 bp in length
            *    96868    96967: gap of unknown length
            *    96968   109418: contig of 12451 bp in length
            *   109419   109518: gap of unknown length
            *   109519   120755: contig of 11237 bp in length
            *   120756   120855: gap of unknown length
            *   120856   134377: contig of 13522 bp in length
            *   134378   134477: gap of unknown length
            *   134478   144432: contig of 9955 bp in length
            *   144433   144532: gap of unknown length
            *   144533   155013: contig of 10481 bp in length
            *   155014   155113: gap of unknown length
            *   155114   161933: contig of 6820 bp in length
            *   161934   162033: gap of unknown length
            *   162034   168622: contig of 6589 bp in length
            *   168623   168722: gap of unknown length
            *   168723   175724: contig of 7002 bp in length
            *   175725   175824: gap of unknown length
            *   175825   180213: contig of 4389 bp in length
            *   180214   180313: gap of unknown length
            *   180314   186640: contig of 6327 bp in length
            *   186641   186740: gap of unknown length
            *   186741   190642: contig of 3902 bp in length
            *   190643   190742: gap of unknown length
            *   190743   194248: contig of 3506 bp in length
            *   194249   194348: gap of unknown length
            *   194349   196413: contig of 2065 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196413
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-11J15"
BASE COUNT    46007 a  51821 c  52327 g  44554 t   1704 others
ORIGIN

  Query Match             39.5%;  Score 394.8;  DB 2;  Length 196413;
  Best Local Similarity   75.8%;  Pred. No. 6.7e-68;
  Matches  547;  Conservative    0;  Mismatches  157;  Indels   18;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC------- 53
              |||||||||||||||||| |||||||||| |||||||||||||| |||||| |
Db     196321 AGTTCAAGACCAGCCTGACCAACATGGTGGAACCCCATCTCTAGTAAAAAAAAAAAAAAA 196262

Qy         54 -------AAAATTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGC 106
                      |||||||| |||| ||||| || ||||| ||||| | |||||| ||||||
Db     196261 AAAAAAAAAAATTAGCCAGGTGTGGTGGGGCATGCCTGAAATCCTAACTACTTGGGAGGC 196202

Qy        107 CGAGGCAGGAGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATG 165
              ||||||||| ||| |||||| | | | ||||||||||||    ||||  | |||| |
Db     196201 TGAGGCAGGAAAATTGCTTGAAGC--CAAGAGGTGGAGGTTGCAATGAGCTGATATCACG 196144

Qy        166 CCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCA 225
              || |||||||||||||  |||||||| | ||||   ||||||| ||||||  |   | |
Db     196143 CCGTTGCACTCCAGCCTGGGCAACAAAAATGAAACTCCATCTCAAAAAAAAGAGGGGCCG 196084

Qy        226 TTTTATATATATATATATATATATACACACACACACACATATATATATACACATATAT 285
                  |  |  ||  |  || | || || | | | | | |||||||||||| | ||||||
Db     196083 GGAGAGGTGGATCAAGCTAGAAATCTATATATATATATATATATATATATATATATAT 196024

Qy        286 ATACACATATATACACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || |
Db     196023 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 195964

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              ||||||||| | |||| |||| | |||||||| | |||||||| | | ||||||| | ||||
Db     195963 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 195904

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| | | ||||||| | ||||||| | ||||||| | ||||| | |||||||| | ||||
Db     195903 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 195844

Qy        466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
              ||| ||||||| | |||||||| ||||| | | |||||||||| | ||||||| | | |||
Db     195843 ATATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATA 195785
```

```
Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | |||||||| | |||||||||| | ||||||| | ||||||| ||||||| | |||
Db     195784 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 195725

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| |||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     195724 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 195665

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||   |||  |
Db     195664 TATATATATATATATATATATATATATATATATATATATATATATATATATATCTATATCCA 195605

Qy        706 TA 707
              ||
Db     195604 CA 195603
```

RESULT 10
AC015468
LOCUS AC015468 189662 bp DNA linear PRI 01-MAY-2001
DEFINITION Homo sapiens chromosome 8, clone RP11-369E15, complete sequence.
ACCESSION AC015468
VERSION AC015468.5 GI:13899433
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 189662)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-369E15
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189662)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 189662)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On May 1, 2001 this sequence version replaced gi:12313808.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L2466
Center clone name: 369_E_15
----------.

FEATURES	Location/Qualifiers
source	1. .189662
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="8"
	/map="8"
	/clone="RP11-369E15"
	/clone_lib="RPCI-11 Human Male BAC"
repeat_region	1. .100
	/rpt_family="MLT1A1"
repeat_region	1594. .1741
	/rpt_family="GA-rich"
repeat_region	complement(2782. .2957)
	/rpt_family="MIR"
repeat_region	3311. .3471
	/rpt_family="MIR"
repeat_region	3964. .4067
	/rpt_family="MIR"
repeat_region	complement(4528. .4771)
	/rpt_family="L2"
repeat_region	4776. .4871
	/rpt_family="MER4A"
repeat_region	4869. .5223
	/rpt_family="MER4A"
repeat_region	complement(5255. .5765)
	/rpt_family="L2"
repeat_region	complement(5805. .6003)
	/rpt_family="THE1A"
repeat_region	6004. .6035
	/rpt_family="(TTA)n"
repeat_region	complement(6151. .6319)
	/rpt_family="SVA"
repeat_region	complement(6320. .6468)
	/rpt_family="THE1A"
repeat_region	8722. .8751
	/rpt_family="(TTTTA)n"
repeat_region	9155. .9553
	/rpt_family="MSTB"
repeat_region	9599. .9982
	/rpt_family="THE1C"
repeat_region	complement(10113. .10320)
	/rpt_family="AluY"
repeat_region	10822. .10970
	/rpt_family="MIR3"
unsure	11210. .11215
	/note="<30 qual SNGL region"
repeat_region	11794. .12080
	/rpt_family="AluSx"
repeat_region	12974. .13010
	/rpt_family="(TG)n"
repeat_region	complement(13091. .13304)
	/rpt_family="MER106A"
repeat_region	13484. .13858
	/rpt_family="MLT1C"
repeat_region	14025. .14260
	/rpt_family="L2"
repeat_region	14584. .14679

/rpt_family="MIR"
repeat_region 14722. .14825
/rpt_family="L2"
repeat_region 15167. .15332
/rpt_family="MER33"
repeat_region 15436. .15788
/rpt_family="L2"
repeat_region complement(16078. .16165)
/rpt_family="AluJ/FRAM"
repeat_region 16532. .17094
/rpt_family="MLT1F1"
repeat_region 17653. .17674
/rpt_family="(TAA)n"
repeat_region complement(17675. .17892)
/rpt_family="MER7C"
repeat_region complement(17897. .18650)
/rpt_family="LTR17"
repeat_region complement(18651. .18996)
/rpt_family="MER7C"
repeat_region 18999. .19103
/rpt_family="(TAA)n"
repeat_region complement(19203. .19283)
/rpt_family="MIR"
repeat_region 20464. .20523
/rpt_family="A-rich"
repeat_region complement(21279. .21902)
/rpt_family="MER41B"
repeat_region 21922. .22092
/rpt_family="MIR"
repeat_region 22249. .22373
/rpt_family="MIR"
repeat_region 23956. .24058
/rpt_family="L2"
repeat_region complement(24069. .24198)
/rpt_family="MIR"
repeat_region 25149. .25213
/rpt_family="L2"
repeat_region complement(25783. .26130)
/rpt_family="MLT1I"
repeat_region complement(26174. .26539)
/rpt_family="MLT1A1"
repeat_region complement(28187. .28528)
/rpt_family="AluY"
repeat_region complement(28792. .29263)
/rpt_family="Charlie8"
repeat_region complement(30652. .30730)
/rpt_family="MER53"
repeat_region complement(30986. .31117)
/rpt_family="L1ME1"
repeat_region 31113. .31556
/rpt_family="L1ME2"
repeat_region 31590. .32174
/rpt_family="L1ME2"
repeat_region 32162. .32588
/rpt_family="L1ME3"
repeat_region 33242. .33490
/rpt_family="MIR"
repeat_region complement(34416. .34803)
/rpt_family="MSTB"
repeat_region complement(35044. .35350)
/rpt_family="AluSg"
repeat_region 36407. .36695
/rpt_family="AluSx"
repeat_region complement(37670. .37790)
/rpt_family="FLAM_A"
repeat_region complement(38065. .38132)
/rpt_family="L2"
repeat_region complement(38504. .38590)
/rpt_family="MIR"
repeat_region 39044. .39228
/rpt_family="MIR3"
repeat_region complement(40317. .40473)
/rpt_family="MIR"

Query Match 39.1%; Score 390.8; DB 9; Length 189662;
Best Local Similarity 65.9%; Pred. No. 4e-67;
Matches 643; Conservative 0; Mismatches 327; Indels 6; Gaps 5;

```
Qy        1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAAT-ACAAAATT 59
            ||||| |||||  ||||  ||||||| ||||||||||||||||||||||| | ||||||
Db    65793 AGTTCGAGACCCACCTGGCCAACATGATGAAACCCCATCTCTACTAAAAATGAAAAAATT 65852

Qy       60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGA-GGCCGAGGCAGGAGA 118
            ||   ||  |||||  |   || | ||||||||||||||  ||| ||| ||||||| |||
Db    65853 AGCTGGGTGTGGTGGCAGGTGCATCTAATCCCAGCTACTCTGGAGGGCTGAGGCAGAAGA 65912

Qy      119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
            ||  ||||||||  | || |||||| ||||| | ||| | |||||  ||||||||||||||
Db    65913 ATTGCTTGAACC--CAGGAGGTGGACGTTGCAGTGAGCCGAGATCGCGCCATTGCACTCC 65970

Qy      178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
            ||||  |||||||||||| |||     ||||| |||||||||          |||||||||
Db    65971 AGCCTGGGCAACAAGAGCAAAACTCTGTCTCAAAAAAAAAAATTTTATAAATATATATAT 66030

Qy      238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
            | |||||||||||| | | | | | | ||||||| ||| | ||| ||||| | | ||||
Db    66031 AAATATATATATAAATATATATAAATATATATATAAATATATATAAATATATATAAATAT 66090

Qy      298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
            | ||||| | | ||||| | | ||||| ||| | ||| |||| || | |||||||||| |
Db    66091 ATATATAAATATATATAAATATATATAAATATATATAAATATATATAAATATATATATAA 66150

Qy      358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
            |||| |||| | ||||| | | ||||||| | | ||||||| | |||||||| | | ||||
Db    66151 ATATATATAAATATATAAAAATATATATATAAATATATATAAATATATATAAATATATAT 66210

Qy      418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
            | | | |||||||| | ||||||| | ||||| | ||||||| | | | ||| |||||||
Db    66211 AAATATATATATAAATATATATAAATATATATAAATATATATATAAAAATATATATATAA 66270

Qy      478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
            | ||||||| |||||| | | | ||||||| | ||||||| | | ||||||| | || ||
Db    66271 ATATATATA-AATATATATATAAATATATATAAATATATATATAAATATATAAATATGTA 66329

Qy      538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
            || | |||||||| | |  |||||| | || |||| | ||||| | |||| |||| | |||
Db    66330 TATAAATATATAAATATGTATATAAATATGTATATAAATATATAAATATGTATATAAATA 66389

Qy      598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
            || | ||| ||| ||| |||    |||||| | |||||||| | | |||||||| | ||  |||
Db    66390 TATATATAAATATATAAATATGTATATAAATATATATATAAATATATATAAATATGTATA 66449

Qy      658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
            || | | | ||| ||||  ||||| ||| | |||||||  ||||||  |  |||||||||
Db    66450 TAAATATATATAAATATGTATATAAATATATATATATAAATATATATAAAAATATATATA 66509

Qy      718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
            ||      | |  |||      | | | ||  |   |||  |  ||| |  | | |
Db    66510 TAAATATATAAATATATATATAAATATATATAAATATATATAAATATATATATAAATATA 66569

Qy      778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
              | |    | ||     | |   | || |       |    |||||       | |
Db    66570 TAAATATATATATATAAATATATATATAAATATATATAAAAATATAAATATATATAAATA 66629

Qy      838 TACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATAT 897
            || ||| |  | ||  | |  |      |  |  ||  |  |    |||| ||||
Db    66630 TATATATAAATATATATATAAAAAATATATAAAAATATATATATAAATATATATAAATAT 66689

Qy      898 ATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAA 957
            |||||   | |      || |  || || |  |   ||  | | |||| || |     ||
Db    66690 ATATAAATATATATAAATATATATATAGTAGTGAAGAATCACTAAAATATTTTTATGAAA 66749

Qy      958 TGGCTAAGATGCCAAA 973
              || ||  | | |
Db    66750 AAAGTATGAATCTAGA 66765
```

```
RESULT 11
AC073310/c
LOCUS       AC073310              125685 bp    DNA     linear   PRI 26-APR-2003
DEFINITION  Homo sapiens BAC clone RP11-49G5 from 7, complete sequence.
ACCESSION   AC073310
VERSION     AC073310.7  GI:13176604
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125685)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 125685)
  AUTHORS   Haakenson,W., Hannah,C. and Kang,K.
  TITLE     The sequence of Homo sapiens BAC clone RP11-49G5
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 125685)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 125685)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 125685)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 125685)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7  (bases 1 to 125685)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Mar 1, 2001 this sequence version replaced gi:11024926.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0049G05
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center. For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-409M7, 200 bp overlap the
            clone sequenced to the right is RP11-374N8.  Actual start of this
            clone is at base position 14572 of RP11-409M7 actual end is at base
            position 125685 of RP11-49G5.
FEATURES             Location/Qualifiers
     source          1. .125685
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP11-49G5"
                     /clone_lib="RPCI-11"
     repeat_region   1013. .1151
                     /rpt_family="CR1"
     variation       1220
                     /allele="G"
                     /allele="T"
                     /db_xref="dbSNP:1525457"
     variation       2173
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:1357936"
     repeat_region   2238. .2394
                     /rpt_family="MIR"
     variation       2366
                     /allele="C"
                     /allele="G"
                     /db_xref="dbSNP:1357943"
     repeat_region   2548. .2668
                     /rpt_family="L2"
     variation       3196
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:1554696"
     repeat_region   3206. .5231
                     /rpt_family="L1"
     variation       3808
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:1554695"
     variation       3956
                     /allele="C"
                     /allele="G"
                     /db_xref="dbSNP:1554694"
     variation       5517
                     /allele="G"
                     /allele="C"
                     /db_xref="dbSNP:955932"
     repeat_region   5694. .5715
                     /rpt_family="AT_rich"
     repeat_region   7142. .7353
                     /rpt_family="MER1_type"
```

```
     repeat_region   7490. .7666
                     /rpt_family="ERVL"
     repeat_region   7671. .7771
                     /rpt_family="L1"
     repeat_region   7951. .7985
                     /rpt_family="AT_rich"
     repeat_region   9572. .9809
                     /rpt_family="L1"
     repeat_region   9811. .9831
                     /rpt_family="AT_rich"
     repeat_region   9895. .10017
                     /rpt_family="MaLR"
     repeat_region   10320. .10375
                     /rpt_family="MIR"
     repeat_region   10395. .10732
                     /rpt_family="L1"
     repeat_region   10911. .11118
                     /rpt_family="(TA)n"
     unsure          10914. .11684
                     /note="unresolved simple sequence repeat"
     repeat_region   11130. .11196
                     /rpt_family="L1"
     repeat_region   11217. .11663
                     /rpt_family="(TA)n"
     repeat_region   11667. .11918
                     /rpt_family="Alu"
     repeat_region   11919. .11970
                     /rpt_family="L1"
     repeat_region   12356. .12508
                     /rpt_family="L1"
     repeat_region   12646. .12813
                     /rpt_family="L1"
     repeat_region   12863. .13111
                     /rpt_family="L1"
     repeat_region   13280. .13592
                     /rpt_family="L1"
     repeat_region   13768. .14032
                     /rpt_family="L1"
     repeat_region   14048. .14489
                     /rpt_family="L1"
     repeat_region   15108. .15378
                     /rpt_family="Alu"
     repeat_region   15563. .15584
                     /rpt_family="AT_rich"
     repeat_region   17280. .17442
                     /rpt_family="MIR"
     repeat_region   18313. .18467
                     /rpt_family="L1"
     repeat_region   18468. .18676
                     /rpt_family="L1"
     repeat_region   18677. .18701
                     /rpt_family="AT_rich"
     repeat_region   18797. .19155
                     /rpt_family="MaLR"
     repeat_region   19594. .19911
                     /rpt_family="Alu"
     variation       20353

  Query Match             38.7%;  Score 387.4;  DB 9;  Length 125685;
  Best Local Similarity   67.7%;  Pred. No. 1.9e-66;
  Matches  647;  Conservative    0;  Mismatches  296;  Indels   12;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||   ||||||||||||||||||| | | ||||||||||||| ||||||
Db      11886 AGTTCAAGACCAGCCTAGCCAACATGGTGAAACCCCGTTTTTACTAAAAATACAAAAATT 11827

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||| ||||| |    ||||||||||||||| ||    ||||||  ||||||||||
Db      11826 AGCCAGGCGTGGTGGT---TGCCTGTAATCCCAGCCACACGGGAGGCTAAGGCAGGAGAA 11770

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              |  |||||| |||  || |  ||||||||| | ||||  |||||| |||| |||||||||
Db      11769 TTGCTTGAATCTG--GGAGACGGAGGTTGCAGTGAGCTGAGATCACGCCACTGCACTCCA 11712

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||  ||  || |||| || |   ||||||| ||||||||||||     | ||||||||||
Db      11711 GCCTGGGTGAC-AGAGTGAGACTCCATCTCAAAAAAAAAAAAAAAAGTCATATATATATA 11653

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              |||||||||||||| | | | | | | | ||||||| | | ||||||| | | |||||
Db      11652 TATATATATATATATATATATATAAAAATATATATATAAATATATATATAAATATA 11593

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              |||||||| | |||| || | ||||||||| | ||| || | || | ||||||| | | |
Db      11592 TATATATAAATATATTTATAAATATATATATAAATATATTTATAAATATATATAAATATA 11533

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||| |||| | ||||||| | ||||||   | | | ||||| | ||| ||| | | |||||
Db      11532 TATATATAAATATATATATAAATATATGTATAAAAATATATATATAAATATAAATATATA 11473

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              || | ||||||| | | | ||| | ||||| | ||| ||| | ||||| | ||||||| |
Db      11472 TATAAATATATATATAAAAATATATATATAAATATAAATATATATATAAATATATATATA 11413

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              |||||||||    ||||| | | ||||| ||| | ||||||| | | ||||| | |  ||
Db      11412 AATATATATATATATATAAATATATATAAATATATATATATATATATATAAATATCTA 11353

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              |||| |||||||||| | ||||||| | | ||| |||||||| | |||||||| | |
Db      11352 TATAAACATATATATATAAATATATATATAAACATATATATATAAATATATATATAAACA 11293

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              |||| | ||| ||| |||||||| |||||||| ||| ||| | | ||||||| | ||| |
Db      11292 TATATATATAAATATATATATAAACATATATATATAAATATATATATATATATATATA 11233

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              |||| | | | ||| | | | |||||||| || |||||||||    | |||||| |||| |
Db      11232 TATATATATATATATAAACATATATATATTCAAATATATATATAAACATATATATATAAA 11173

Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775
              |||||     |   | |||          | |    ||   |       |   | | |
Db      11172 TATATATATAAACATATATATATAAATATATATATAAACATATATAATATAACATATATA 11113

Qy        776 GAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGA 835
              | |  | | |    | ||   | |  | || |     | |    | |||       |
Db      11112 TATATATATATATATATAAATATATATATAAATATATATATAAATATATATATATAAATA 11053

Qy        836 AATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATA 894
              ||| |||||| | ||   | |  ||| | || |  ||  ||    |   |  ||||||
Db      11052 TATATATAAATATATATATATATATATATATAAATATATATATATAAATATATATATA 10993

Qy        895 TATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTA 949
              || |||||   | |    | || |  ||||          ||  || || || | |
Db      10992 TAAATATATATATATATATATATATATAAATATATATATAAATATATATATATAA 10938


RESULT 12
AL133477
LOCUS       AL133477               139118 bp    DNA     linear   PRI 02-FEB-2001
DEFINITION  Human DNA sequence from clone RP11-172F4 on chromosome 9q22.2-31.1,
            complete sequence.
ACCESSION   AL133477
VERSION     AL133477.16  GI:12666198
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 139118)
  AUTHORS   Lloyd,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 5, 2001 this sequence version replaced gi:12584335.
```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr9
RP11-172F4 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-172F4 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP11-172F4 is at 1 in this sequence. The true left end of clone RP11-240L7 is at 139019 in this sequence. The true right end of clone RP11-392G7 is at 43511 in this sequence.

FEATURES Location/Qualifiers

source	1. .139118
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="9"
	/map="q22.2-31.1"
	/clone="RP11-172F4"
	/clone_lib="RPCI-11.1"
repeat_region	45. .78
	/note="L1MB2 repeat: matches 6125. .6158 of consensus"
repeat_region	195. .361
	/note="MER20 repeat: matches 39. .218 of consensus"
repeat_region	488. .688
	/note="MLT1B repeat: matches 205. .390 of consensus"
repeat_region	689. .986
	/note="AluY repeat: matches 1. .297 of consensus"
repeat_region	987. .1144
	/note="MLT1B repeat: matches 5. .205 of consensus"
repeat_region	1149. .1500
	/note="MST-INTERNAL repeat: matches 1289. .1651 of consensus"
repeat_region	1812. .2130
	/note="MLT1-INTERNAL repeat: matches 404. .728 of consensus"
repeat_region	2175. .2472
	/note="MLT2FB repeat: matches 101. .414 of consensus"
repeat_region	2729. .2832
	/note="MER5B repeat: matches 2. .80 of consensus"
repeat_region	2833. .3146
	/note="AluSx repeat: matches 1. .308 of consensus"
repeat_region	3147. .3236
	/note="MER5B repeat: matches 80. .177 of consensus"
repeat_region	5694. .6008
	/note="AluSx repeat: matches 1. .306 of consensus"
repeat_region	6183. .6318
	/note="MER5A repeat: matches 10. .189 of consensus"
repeat_region	6556. .6717
	/note="FRAM repeat: matches -5. .161 of consensus"
repeat_region	7313. .7619
	/note="AluSx repeat: matches 1. .311 of consensus"
repeat_region	7907. .8217
	/note="AluY repeat: matches 1. .311 of consensus"
repeat_region	8545. .8837
	/note="AluSg repeat: matches 1. .293 of consensus"
repeat_region	9101. .9389
	/note="AluJo repeat: matches 1. .282 of consensus"
repeat_region	9399. .9701
	/note="AluY repeat: matches 1. .311 of consensus"
repeat_region	9702. .10014
	/note="AluSx repeat: matches 1. .311 of consensus"
repeat_region	10050. .10292
	/note="AluSp repeat: matches 66. .310 of consensus"
repeat_region	10300. .10615
	/note="AluSx repeat: matches 4. .312 of consensus"
repeat_region	14460. .14754
	/note="AluJo repeat: matches 13. .308 of consensus"
repeat_region	14846. .15042
	/note="AluJb repeat: matches 96. .296 of consensus"
repeat_region	15166. .15281
	/note="AluJb repeat: matches 9. .96 of consensus"
repeat_region	15506. .15781
	/note="AluSg repeat: matches 1. .276 of consensus"
repeat_region	16165. .16466
	/note="AluSx repeat: matches 1. .300 of consensus"
repeat_region	18757. .19057
	/note="AluSx repeat: matches 1. .301 of consensus"
repeat_region	19097. .19144
	/note="12 copies 4 mer tata 77% conserved"
repeat_region	19359. .19662
	/note="AluSg repeat: matches 1. .304 of consensus"
repeat_region	19743. .20060
	/note="AluSx repeat: matches 1. .305 of consensus"
repeat_region	21872. .22179
	/note="AluSx repeat: matches 1. .299 of consensus"
repeat_region	27049. .27427
	/note="MLT1B repeat: matches 7. .390 of consensus"
repeat_region	29141. .29384
	/note="MIR repeat: matches 2. .260 of consensus"
misc_feature	29833. .30559
	/note="CpG island"
	/evidence=not_experimental
repeat_region	31569. .31653
	/note="MER91A repeat: matches 3. .63 of consensus"
repeat_region	31939. .32227
	/note="AluY repeat: matches 1. .291 of consensus"
repeat_region	32228. .32323
	/note="48 copies 2 mer gg 67% conserved"
repeat_region	32248. .32297
	/note="10 copies 5 mer gacgg 88% conserved"
repeat_region	32324. .32378
	/note="11 copies 5 mer aagaa 78% conserved"
repeat_region	33239. .33518
	/note="AluJo repeat: matches 1. .289 of consensus"
repeat_region	34313. .34610
	/note="AluSq repeat: matches 1. .298 of consensus"
repeat_region	35793. .35899
	/note="MIR repeat: matches 1. .116 of consensus"
repeat_region	37385. .37604
	/note="L2 repeat: matches 2436. .2659 of consensus"
repeat_region	38070. .38181
	/note="L2 repeat: matches 2563. .2729 of consensus"
repeat_region	40862. .41123
	/note="AluSq repeat: matches 39. .302 of consensus"
repeat_region	41722. .42016
	/note="AluSp repeat: matches 1. .297 of consensus"
repeat_region	43048. .43521
	/note="L1MB5 repeat: matches 4783. .5239 of consensus"
repeat_region	43522. .43695
	/note="FAM repeat: matches 1. .174 of consensus"
repeat_region	43696. .44552
	/note="L1MB5 repeat: matches 5239. .6138 of consensus"
repeat_region	44598. .44834
	/note="L1MB4 repeat: matches 5814. .6065 of consensus"
repeat_region	44835. .45133

```
                     /note="AluSg repeat: matches 1. .295 of consensus"
     repeat_region   45134. .45257
                     /note="L1MB4 repeat: matches 6065. .6184 of consensus"
     repeat_region   45312. .45611
                     /note="AluJo repeat: matches 1. .293 of consensus"
     repeat_region   45681. .45989
                     /note="AluSg repeat: matches 1. .300 of consensus"
     repeat_region   46204. .46347
                     /note="L1ME2 repeat: matches 6007. .6153 of consensus"
     repeat_region   46876. .47188
                     /note="AluJo repeat: matches 1. .304 of consensus"
     repeat_region   47262. .47560
                     /note="AluSq repeat: matches 1. .301 of consensus"
     repeat_region   47923. .48041
                     /note="L1MC4 repeat: matches 6562. .6679 of consensus"
     repeat_region   48060. .48148
                     /note="L1ME1 repeat: matches 5347. .5428 of consensus"
     repeat_region   48149. .48413
                     /note="AluJo repeat: matches 2. .302 of consensus"
     repeat_region   48414. .48641
                     /note="L1ME1 repeat: matches 5428. .5637 of consensus"
     repeat_region   48642. .48935
                     /note="AluSg repeat: matches 1. .294 of consensus"
     repeat_region   48936. .49372
                     /note="L1ME1 repeat: matches 5637. .6160 of consensus"
     repeat_region   50733. .50947
                     /note="MIR repeat: matches 2. .257 of consensus"
     repeat_region   51470. .51541
                     /note="MIR repeat: matches 44. .124 of consensus"
     repeat_region   51667. .51977
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   52966. .53214
                     /note="MIR repeat: matches 12. .262 of consensus"
     repeat_region   53551. .54965
                     /note="L2 repeat: matches 505. .2101 of consensus"
     repeat_region   54984. .55272
                     /note="AluSp repeat: matches 1. .291 of consensus"

  Query Match           38.6%;  Score 385.8;  DB 9;  Length 139118;
  Best Local Similarity   77.4%;  Pred. No. 3.8e-66;
  Matches  583;  Conservative    0;  Mismatches  132;  Indels   38;  Gaps    8;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57
              ||||| ||||||||||| ||||||||||||||||||||||||||||||||||   ||||
Db      97498 AGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAAAA 97557

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              ||||   ||| |||||  ||| ||||||| || ||||||| | |||||| ||||||||||
Db      97558 TTAGCTGGGCGTGGTGGCACATGCCTGTAGTCTCAGCTACCTGGGAGGCTGAGGCAGGAG 97617

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTC 176
              |||| ||||||||  | || ||||||||||||| | ||| | |||||   ||| |||||||
Db      97618 AATCGCTTGAACC--CAGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCTCCACTGCACTC 97675

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCAT---------- 226
              | ||   ||| || |||||   |   | ||||| ||||||||||   ||
Db      97676 CGGCTTGGGCGAC-AGAGCATGACTCCGTCTCAAAAAAAAAAAAATATATATATATATAT 97734

Qy        227 --------------TTTATATATATATATATATATATACACACACACACATATAT 271
                           | |||||||||| | ||||||| ||| | ||| ||| | ||| ||
Db      97735 ATGTGTGTGTGTGTGTATATATATACACATATATACATATATACATACATATATACAT 97794

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331
              ||||||| | |||||||||||||||||||| ||| ||||||||||||| ||||| | ||
Db      97795 ATATACATACATATATACACATATATACATACATATACATATATACACACATATACATAC 97854

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391
              ||| | || || ||||||||| ||||| ||  ||| | | ||||| ||||||||||||
Db      97855 ATATACATATATACATATATACACACACATACATATATACATATATACATATATACACAT 97914

Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451
              ||||| | | ||| ||| | ||| ||||| ||| | ||| ||||| ||| |||||||| ||
Db      97915 ATATACATATATACATATATACACATATATACACACATATACATACATATACATATATAC 97974

Qy        452 ATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATACAC 510
              |||||||||||||||||||||||||||||| ||||| || ||||||| | ||| |||||||
Db      97975 GTATATACACATATATACATATATACATATATACGTATATATACATATATACATATATAC 98034

Qy        511 ATATATACACACATATATACACATATATACACATATATATACACATATATACACATATAT 570
              |||||| ||| ||||| | | |||  || | ||| ||||| |||||||| ||||||||
Db      98035 GTATATATACATATATACATATATACGTATATATACATATATACATATATACATATAT 98094

Qy        571 ACATATATACACATAT----ATATACATATATACACATATATACATATATACACATATAT 626
              |||||||| ||||||    ||||| | ||||| || ||||| |||||||| ||||||||
Db      98095 ACATATATATACATATATACATATATACATATATACGTATATGCATATATATACATATAT 98154

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              ||||||||| ||| ||||||||| ||| | | | | ||| | |||||||||| ||||||
Db      98155 ACATATATATACATATATATACATATATACACATATATACACA-ATATACATACATATAC 98213

Qy        687 ACATATATATACGTATATATATGTATATATATA 719
              | ||||| | |  || |||||| || ||||||||
Db      98214 ATATATACACATATACATATATATACATATATA 98246
```

```
RESULT 13
HUAC004787
LOCUS       HUAC004787            216021 bp    DNA     linear   PRI 30-OCT-2002
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10, complete
            sequence.
ACCESSION   AC004787
VERSION     AC004787.1  GI:3337381
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 216021)
  AUTHORS   Loftus,B.J., Kim,U.J., Sneddon,V.P., Kalush,F., Brandon,R.,
            Fuhrmann,J., Mason,T., Crosby,M.L., Barnstead,M., Cronin,L.,
            Deslattes Mays,A., Cao,Y., Xu,R.X., Kang,H.L., Mitchell,S.,
            Eichler,E.E., Harris,P.C., Venter,J.C. and Adams,M.D.
  TITLE     Genome duplications and other features in 12 Mb of DNA sequence
            from human chromosome 16p and 16q
  JOURNAL   Genomics 60 (3), 295-308 (1999)
  MEDLINE   99425270
   PUBMED   10493829
REFERENCE   2  (bases 1 to 216021)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, Email:
            bjloftus@tigr.org
REFERENCE   4  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jul 24, 1998 this sequence version replaced gi:3241936.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org.   The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
```

```
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES             Location/Qualifiers
     source          1. .216021
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="#16q21-22"
                     /clone="A-952F10"
     STS             27765. .27872
                     /note="7766, STS1-cSRL-27g3-uA/cSRL-27g3-uZ, Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02280"
     STS             73826. .73943
                     /note="7608, STS1-cSRL-24g1-uA/cSRL-24g1-uZ, Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02122"
     STS             175801. .175945
                     /note="16084, CHLC.GCT10B02, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G09703"
     STS             175810. .175945
                     /note="16316, CHLC.GCT15C04, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G09935"
     STS             199463. .199572
                     /note="9824, WI-3555, Chr. 16, Homo sapiens"
                     /db_xref="dbSTS:G04338"
BASE COUNT    60960 a  51778 c  49172 g  53987 t    124 others
ORIGIN

  Query Match           38.3%;  Score 382.6;  DB 9;  Length 216021;
  Best Local Similarity   76.2%;  Pred. No. 1.6e-65;
  Matches  551;  Conservative    0;  Mismatches  159;  Indels   13;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||| | |||| ||||  |||||| |||||||||||||||||||||||    |||||||
Db      80976 AGTTCGAAACCACCCTGGCCAACATAGTGAAACCCCATCTCTACTAAAATACAAAAATTA 81035

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              | ||||| |||||  | |||||||||||||||||||||  |||||| |||||||||| ||
Db      81036 GCCAGGCGTGGTGGCAGACGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGGAT 81095

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCAG 179
               |||||||  | || ||  |||||||| | ||| | ||||  ||||| || |||||||
Db      81096 TGCTTGAACC--CAGGAGGCAGAGGTTGCAGTGAGCCCAGATTGTGCCACTGGACTCCAG 81153

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || ||||  | |   | ||||| | | | | | |       || | | ||||
Db      81154 CCTGGGAGAC-AGAGTAAGACTCCGTCTCAAATACATATATA-----CATACACACATAT 81207

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||| | ||||||| ||||||| | ||||| | ||| ||||| | | | ||||| | ||
Db      81208 ATATACACATATACATACACACATATATATACACATATACATACACACATACATACACAC 81267

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              |||||||||||| ||||||| ||||||| ||| ||| |||  || ||| ||||||||||||
Db      81268 ATATATACACACATATACATATATATACACATATATATACATATACACATATATACACAT 81327

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || || | | | | | | | ||||| | | | | | ||| ||| ||||| ||| |||||
Db      81328 ATATA-ATATACACACATATATATACATATATACACATATACACATATATACATATATAC 81386

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||||||||||||||||||||||||| ||| |||||||||||| ||| | |||||| ||||
Db      81387 ACACATATATACACATATATACATATACACACATATATACATATACACACATATACACAC 81446

Qy        480 ATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATAT 536
              ||| | | |   | | ||||||||| | | | ||| | | | ||||| | | | ||| |
Db      81447 ATACACACACATACACACACACATACACACATACACACACATACACACACACATACACAC 81506

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              | | ||| | ||| ||||| ||||||||| ||| |  || | | ||| | ||||||| ||
Db      81507 ACATACACACATACACACACATATACACACATACATGTACACATACACACATATACACAT 81566

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656
              ||||||| |||||||||||||||| ||||||||||||||||||||||||| ||||| ||
Db      81567 ATACACACATATACATATATACACACATATACATATATACACACATATATATACATATAT 81626

Qy        657 ATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATAT 716
              | ||| | | | | ||||||| |||||||| ||||  | |  |||| |||| || | ||
Db      81627 ACACATATATACATATATACACATATATACATATATGCACATATATACATATATACACAT 81686

Qy        717 ATA 719
               |||
Db      81687 GTA 81689


RESULT 14
AY007685
LOCUS       AY007685               92564 bp    DNA     linear   PRI 22-FEB-2002
DEFINITION  Homo sapiens telomerase catalytic subunit (TERT) and sodium
            channel-like protein genes, complete cds.
ACCESSION   AY007685
VERSION     AY007685.2  GI:15991796
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 92564)
  AUTHORS   Leem,S.H., Londono-Vallejo,J.A., Kim,J.H., Bui,H., Tubacher,E.,
            Solomon,G., Park,J.E., Horikawa,I., Kouprina,N., Barrett,J.C. and
            Larionov,V.
  TITLE     The human telomerase gene: complete genomic sequence and analysis
            of tandem repeat polymorphisms in intronic regions
  JOURNAL   Oncogene 21 (5), 769-777 (2002)
  MEDLINE   21839826
   PUBMED   11850805
REFERENCE   2  (bases 1 to 92564)
  AUTHORS   Londono-Vallejo,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-2000) Centre d'Etudes du Polymorphisme Humain, 27
            rue Juliette Dodu, Paris 75010, France
REFERENCE   3  (bases 1 to 92564)
  AUTHORS   Londono-Vallejo,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-2001) Centre d'Etudes du Polymorphisme Humain, 27
            rue Juliette Dodu, Paris 75010, France
  REMARK    Sequence update by submitter
COMMENT     On Oct 10, 2001 this sequence version replaced gi:12642956.
FEATURES             Location/Qualifiers
     source          1. .92564
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5p13"
     repeat_region   complement(2853. .3034)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(6174. .6429)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(7445. .7837)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(8199. .8341)
                     /rpt_family="MLT1"
                     /rpt_type=dispersed
     repeat_region   complement(10300. .10916)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   12028. .12301
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   13088. .13597
                     /rpt_family="L1"
```

```
                /rpt_type=dispersed
repeat_region   13716. .13988
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   14271. .14556
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   14599. .17711
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   18109. .18398
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(20818. .21123)
                /rpt_family="LTR1"
                /rpt_type=dispersed
repeat_region   21650. .21912
                /rpt_family="MLT1"
                /rpt_type=dispersed
repeat_region   complement(22236. .22484)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(22986. .23283)
                /rpt_family="Alu"
                /rpt_type=dispersed
gene            <24442. .64466
                /gene="TERT"
mRNA            join(<24442. .24660,24765. .26118,35473. .35668,37759. .37939,
                38627. .38806,39301. .39456,45360. .45455,46436. .46521,
                49007. .49120,51105. .51176,53048. .53236,57040. .57166,
                58981. .59042,62229. .62353,63135. .63272,63809. .>63912)
                /gene="TERT"
                /product="telomerase catalytic subunit"
CDS             join(24442. .24660,24765. .26118,35473. .35668,37759. .37939,
                38627. .38806,39301. .39456,45360. .45455,46436. .46521,
                49007. .49120,51105. .51176,53048. .53236,57040. .57166,
                58981. .59042,62229. .62353,63135. .63272,63809. .63912)
                /gene="TERT"
                /codon_start=1
                /product="telomerase catalytic subunit"
                /protein_id="AAG23289.1"
                /db_xref="GI:12642957"
                /translation="MPRAPRCRAVRSLLRSHYREVLPLATFVRRLGPQGWRLVQRGDP
                AAFRALVAQCLVCVPWDARPPPAAPSFRQVSCLKELVARVLQRLCERGAKNVLAFGFA
                LLDGARGGPPEAFTTSVRSYLPNTVTDALRGSGAWGLLLRRVGDDVLVHLLARCALFV
                LVAPSCAYQVCGPPLYQLGAATQARPPPHASGPRRRLGCERAWNHSVREAGVPLGLPA
                PGARRRGGSASRSLPLPKRPRRGAAPEPERTPVGQGSWAHPGRTRGPSDRGFCVVSPA
                RPAEEATSLEGALSGTRHSHPSVGRQHHAGPPSTSRPPRPWDTPCPPVYAETKHFLYS
                SGDKEQLRPSFLLSSLRPSLTGARRLVETIFLGSRPWMPGTPRRLPRLPQRYWQMRPL
                FLELLGNHAQCPYGVLLKTHCPLRAAVTPAAGVCAREKPQGSVAAPEEEDTDPRRLVQ
                LLRQHSSPWQVYGFVRACLRRLVPPGLWGSRHNERRFLRNTKKFISLGKHAKLSLQEL
                TWKMSVRDCAWLRRSPGVGCVPAAEHRLREEILAKFLHWLMSVYVVELLRSFFYVTET
                TFQKNRLFFYRKSVWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPALLTSRLRFI
                PKPDGLRPIVNMDYVVGARTFRREKRAERLTSRVKALFSVLNYERARRPGLLGASVLG
                LDDIHRAWRTFVLRVRAQDPPPELYFVKVDVTGAYDTIPQDRLTEVIASIIKPQNTYC
                VRRYAVVQKAAHGHVRKAFKSHVSTLTDLQPYMRQFVAHLQETSPLRDAVVIEQSSSL
                NEASSGLFDVFLRFMCHHAVRIRGKSYVQCQGIPQGSILSTLLCSLCYGDMENKLFAG
                IRRDGLLLRLVDDFLLVTPHLTHAKTFLRTLVRGVPEYGCVVNLRKTVVNFPVEDEAL
                GGTAFVQMPAHGLFPWCGLLLDTRTLEVQSDYSSYARTSIRASLTFNRGFKAGRNMRR
                KLFGVLRLKCHSLFLDLQVNSLQTVCTNIYKILLLQAYRFHACVLQLPFHQQVWKNPT
                FFLRVISDTASLCYSILKAKNAGMSLGAKGAAGPLPSEAVQWLCHQAFLLKLTRHRVT
                YVPLLGSLRTAQTQLSRKLPGTTLTALEAAANPALPSDFKTILD"
repeat_region   27341. .27628
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   27650. .29650
                /note="VNTR"
                /rpt_type=tandem
                /rpt_unit=gagtgaggcgccgtccccgggtgtccctgtcacgtgcagggt
repeat_region   complement(31926. .32203)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(32555. .32835)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   33569. .33861
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(33873. .33967)
                /rpt_family="MER33"
                /rpt_type=dispersed
repeat_region   34111. .35453
                /note="VNTR"
                /rpt_type=tandem
repeat_region   complement(35993. .36281)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   41485. .42324
                /note="VNTR"
                /rpt_type=tandem
                /rpt_unit=gtgggattggttttcatgtgtggggtaggtggggatct
repeat_region   complement(42656. .42954)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   43312. .43618
                /rpt_family="MER1"
                /rpt_type=dispersed
repeat_region   44225. .45115
                /note="VNTR"
                /rpt_type=tandem
                /rpt_unit=ggggtctgatgtgtggtgactgtggatggcggtcgt
repeat_region   complement(48015. .48300)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(49647. .50276)
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   53945. .54234
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   57630. .58735
                /note="VNTR"
                /rpt_type=tandem
                /rpt_unit=tgggcatctgcgtccactccccctctcctg
repeat_region   61332. .61597
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(66201. .66323)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   66389. .66661
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   66661. .66736
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit=aag
repeat_region   67140. .67583
                /rpt_family="MLT1"
                /rpt_type=dispersed
mRNA            complement(join(<73264. .73377,73925. .74084,76991. .77080,
                78049. .78146,79551. .79661,81950. .82131,84724. .84861,
                85253. .85393,91909. .>92068))
                /product="sodium channel-like protein"
CDS             complement(join(73264. .73377,73925. .74084,76991. .77080,
                78049. .78146,79551. .79661,81950. .82131,84724. .84861,
                85253. .85393,91909. .92068))
                /codon_start=1
                /evidence=not_experimental
                /product="sodium channel-like protein"
                /protein_id="AAG23290.1"
                /db_xref="GI:12642958"

Query Match             38.0%;  Score 379.8;  DB 9;  Length 92564;
Best Local Similarity   75.0%;  Pred. No. 5.7e-65;
Matches  542;  Conservative    0;  Mismatches  172;  Indels    9;  Gaps    5;
```

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| ||||||||||  ||| || | |||||||||||| |||||||| || ||||||
Db      11380 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 11439

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db      11440 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 11499

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||      |   ||||||    ||| ||||||||||
Db      11500 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 11554

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db      11555 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 11613

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db      11614 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 11673

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATAT-ATACACA 358
              |||||||| ||||| ||| ||||| ||   | |||||||| || |||||||| ||| |
Db      11674 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 11733

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||  |||| | |||||||||| ||||| | | |||||||||| | | ||||| | | ||| |
Db      11734 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 11793

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              || | | ||||| |||||||||| | ||| | | ||| ||||| ||||| | ||| ||| |
Db      11794 TATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACATATA 11853

Qy        479 CATATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||||||| ||||| | | ||| ||||| | |||||||||| | ||| ||| |||||||
Db      11854 TAAATATATACATATATAAATATACATATATAAATATATACATATATAAATATACATATA 11913

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |||| ||||| ||| ||||||| | | ||| ||| ||||| ||| ||||| ||| | |||
Db      11914 TACATATATAAATATACATATACATATATAAATATATATACACATATATAAATATACATA 11973

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || | ||||| | ||||| | | ||| ||| ||| ||| ||| ||||| | | ||| | |
Db      11974 TATAAATATACATATATAAATATATACATATATAAATATACATATATAAATATATATACA 12033

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | | ||| | ||| | |  |||| ||| |||||||| |||||| |||||||||
Db      12034 TATATAAATATATAAATATACAAGTATATACAAATATATACATATATAAATGTATATACG 12093

Qy        718 TAT 720
              |||
Db      12094 TAT 12096
```

```
RESULT 15
AC114955/c
LOCUS       AC114955               161252 bp    DNA     linear   HTG 14-MAR-2002
DEFINITION  Homo sapiens chromosome 5 clone RP11-117B23, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC114955
VERSION     AC114955.1  GI:19424416
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161252)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161252)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 451340
            Center clone name: RPCI-11_117B23
            -------
            Summary Statistics
            Consensus quality: 148939 bases at least Q40
            Consensus quality: 156608 bases at least Q30
            Consensus quality: 158826 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 160452; sum-of-contigs estimation
            Quality coverage:  7.26 in Q20 bases; agarose-fp estimation
            Quality coverage:  7.92 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3008: contig of 3008 bp in length
            *     3009     3108: gap of unknown length
            *     3109     9209: contig of 6101 bp in length
            *     9210     9309: gap of unknown length
            *     9310    15613: contig of 6304 bp in length
            *    15614    15713: gap of unknown length
            *    15714    23243: contig of 7530 bp in length
            *    23244    23343: gap of unknown length
            *    23344    39674: contig of 16331 bp in length
            *    39675    39774: gap of unknown length
            *    39775    60001: contig of 20227 bp in length
            *    60002    60101: gap of unknown length
            *    60102    85303: contig of 25202 bp in length
            *    85304    85403: gap of unknown length
            *    85404   117460: contig of 32057 bp in length
            *   117461   117560: gap of unknown length
            *   117561   161252: contig of 43692 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161252
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-117B23"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    36040 a  45224 c  43948 g  35239 t    801 others
ORIGIN
```

```
  Query Match           37.9%;  Score 379.4;  DB 2;  Length 161252;
  Best Local Similarity  68.7%;  Pred. No. 6.7e-65;
  Matches  600;  Conservative    0;  Mismatches  256;  Indels   17;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| ||||||||||  ||| || | |||||||||||| |||||||| || ||||||
Db     149753 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 149694

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db     149693 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 149634

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||      |   ||||||    ||| ||||||||||
Db     149633 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 149579

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
```

```
Db     149578 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 149520

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db     149519 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 149460

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATA----- 354
              ||||||||| ||||| ||| ||||| ||   | |||||||| || | | |||||||
Db     149459 ATATATACATATATACACATATATAAATCTATATACATATATATATACATATATAATATA 149400

Qy        355 --CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
                ||||||  || | | ||| ||| | | |||||||  | ||| ||||| ||||| | | |
Db     149399 TTTACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATA 149340

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              |||||||| | ||||| | | ||| ||| ||||||| | ||| ||| | ||| |||||||
Db     149339 CATATATAAATATATACATATATAAATATACATATATAAATACATATATATAAATACATA 149280

Qy        473 TATACACATATATAT---AATATACACACATATATATACACATATATACACACATATATA 529
              |||| | ||| ||||   |||||| ||| ||||| ||| ||||||||| | | ||| |||
Db     149279 TATAAATATACATATATAAATATATACATATATAAATATACATATATAAATATATACATA 149220

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | | ||||||| ||||| ||| | | ||||||| ||| ||| ||| ||| | ||| | |
Db     149219 TATAAATATACATATATACATATATAAATATACATATACATATATAAATATATATACACA 149160

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              || ||| ||| |||||||||| ||||| | | ||| ||| ||| ||| | | ||||| | |
Db     149159 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 149100

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              ||| |||||  | ||| ||| | ||| ||| |||||||  |||||| |  |||||| ||
Db     149099 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 149040

Qy        710 TATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGG 769
              |||| | ||||  |    | | ||       | |  |  ||    |  ||| | |
Db     149039 TATAAACATATATACGTATATACGTATATACATGTATATACGTATATACGTATATACATG 148980

Qy        770 AAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGT 829
              | | |    |  ||     |  ||    | | | ||  |     ||     |  ||
Db     148979 TATATACGTATATACGTATATACATGTATATACGTATATATATAAATGTATATACGTATA 148920

Qy        830 AGGGGAAATACATAAATGTGTAATAAAACCTAT 862
                | ||| |||||| | ||| ||||| | |
Db     148919 TACATATATATATAAATATATAAAAAAACTTTT 148887
```

Search completed: November 7, 2003, 11:11:04
Job time : 3927.96 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 01:54:16 ; Search time 7837.84 Seconds
(without alignments)
10444.222 Million cell updates/sec

Title: US-09-939-209A-3_COPY_3000_5000
Perfect score: 2001
Sequence: 1 atttgtataattagcacctc.....cagttcttctaaagcctatt 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2001	100.0	20300	6	AX451337	AX451337 Sequence
	2	1999.4	99.9	165329	9	AL583850	AL583850 Human DNA
c	3	1991.4	99.5	191699	2	AC031977	AC031977 Homo sapi
c	4	231.2	11.6	234560	2	AC111539	AC111539 Rattus no
c	5	231.2	11.6	275631	2	AC125563	AC125563 Rattus no
c	6	196.4	9.8	167891	9	AC013439	AC013439 Homo sapi
c	7	192	9.6	90429	9	HS55I10	AL035408 Human DNA
	8	192	9.6	110000	2	AC120598_2	Continuation (3 of
	9	192	9.6	117000	9	AC069222	AC069222 Homo sapi
c	10	192	9.6	140952	2	AC027182	AC027182 Homo sapi
	11	192	9.6	166992	2	AC022883	AC022883 Homo sapi
c	12	192	9.6	204158	9	AL133383	AL133383 Human DNA
c	13	191.6	9.6	175447	9	AC091607	AC091607 Homo sapi
	14	191.6	9.6	176417	2	AC024160	AC024160 Homo sapi
c	15	191	9.5	134995	9	AL592156	AL592156 Human DNA
	16	190.4	9.5	166087	2	AC022036	AC022036 Homo sapi
	17	190.4	9.5	172641	9	AC084235	AC084235 Homo sapi
c	18	190	9.5	74673	2	AC027814	AC027814 Homo sapi
c	19	186.8	9.3	188208	9	AL441986	AL441986 Human DNA
c	20	186.8	9.3	193784	2	AL606964	AL606964 Homo sapi
	21	186.8	9.3	193787	9	AC017111	AC017111 Homo sapi
c	22	186	9.3	167303	9	AP001527	AP001527 Homo sapi
c	23	186	9.3	169513	9	AP002777	AP002777 Homo sapi
c	24	186	9.3	189637	2	AC009765	AC009765 Homo sapi
c	25	186	9.3	192104	9	AC006994	AC006994 Homo sapi
	26	185.8	9.3	87845	9	AL583860	AL583860 Human DNA
	27	185.2	9.3	60532	9	AL590672	AL590672 Human DNA
c	28	185.2	9.3	129355	9	AC004853	AC004853 Homo sapi
	29	185.2	9.3	170543	9	AC021012	AC021012 Homo sapi
c	30	185.2	9.3	170544	9	CNS05TCC	AL355093 Human chr
	31	185.2	9.3	291371	2	AL355584	AL355584 Homo sapi
	32	184.6	9.2	200807	9	AC073278	AC073278 Homo sapi
	33	184.2	9.2	211791	9	AL162430	AL162430 Human DNA
c	34	183.2	9.2	97811	9	AC008634	AC008634 Homo sapi
c	35	183.2	9.2	115915	9	AC105922	AC105922 Homo sapi
	36	182.4	9.1	197168	2	AC055765	AC055765 Homo sapi
	37	182.2	9.1	149731	9	HS162O13	AL035552 Human DNA
	38	181.6	9.1	55469	9	AL445490	AL445490 Human DNA
c	39	181	9.0	121848	9	AC109456	AC109456 Homo sapi
	40	181	9.0	206691	9	AC018752	AC018752 Homo sapi
c	41	180.6	9.0	132718	9	AC107908	AC107908 Homo sapi
c	42	180.4	9.0	118097	9	AL589826	AL589826 Human DNA
c	43	180.2	9.0	62871	9	AL356860	AL356860 Human DNA
	44	180.2	9.0	128277	2	AC069107	AC069107 Homo sapi
c	45	180.2	9.0	137127	2	AC069194	AC069194 Homo sapi

ALIGNMENTS

RESULT 1
AX451337
LOCUS AX451337 20300 bp DNA linear PAT 03-JUL-2002
DEFINITION Sequence 3 from Patent WO0216653.
ACCESSION AX451337
VERSION AX451337.1 GI:21698388
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Levitt,P.R., Mirnics,K., Kodavali,V.C. and Nimgaonkar,V.L.
TITLE Methods and systems for facilitating the diagnosis and treatment of schizophrenia
JOURNAL Patent: WO 0216653-A 3 28-FEB-2002;

```
          University of Pittsburgh (US)
FEATURES             Location/Qualifiers
     source          1. .20300
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="A genomic sequence containing RGS4 nucleic acid
                     sequence and sequ ences upstream and downstream to the
                     RGS4 nucleic acid sequence"
BASE COUNT     6157 a   4102 c   3775 g   6266 t
ORIGIN

  Query Match             100.0%;  Score 2001;  DB 6;  Length 20300;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2001;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3000 ATTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAG 3059

Qy         61 AGAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3060 AGAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAG 3119

Qy        121 TCATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3120 TCATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGT 3179

Qy        181 CATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3180 CATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAA 3239

Qy        241 TATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3240 TATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGC 3299

Qy        301 AGGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3300 AGGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTA 3359

Qy        361 TAATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3360 TAATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAAT 3419

Qy        421 AGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3420 AGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTA 3479

Qy        481 CATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3480 CATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTC 3539

Qy        541 ACTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3540 ACTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTG 3599

Qy        601 CCCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3600 CCCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTAC 3659

Qy        661 CCACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3660 CCACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACC 3719

Qy        721 TTTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3720 TTTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCA 3779

Qy        781 AATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3780 AATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACT 3839

Qy        841 TCTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3840 TCTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTG 3899

Qy        901 ACCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3900 ACCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAA 3959

Qy        961 TGTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3960 TGTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAG 4019

Qy       1021 TGGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4020 TGGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATG 4079

Qy       1081 CTAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAAC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4080 CTAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAAC 4139

Qy       1141 TAGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4140 TAGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAAT 4199

Qy       1201 CCAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4200 CCAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCA 4259

Qy       1261 CCTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4260 CCTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGC 4319

Qy       1321 AGAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4320 AGAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAA 4379

Qy       1381 ACAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4380 ACAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGA 4439

Qy       1441 TGCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4440 TGCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCA 4499

Qy       1501 CAAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4500 CAAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCC 4559

Qy       1561 TGTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4560 TGTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAA 4619

Qy       1621 AATCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4620 AATCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGG 4679

Qy       1681 GAGGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4680 GAGGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGC 4739

Qy       1741 TGAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4740 TGAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAG 4799

Qy       1801 AACTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4800 AACTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATC 4859

Qy       1861 CTTGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4860 CTTGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACC 4919

Qy       1921 TAGAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCCCAGTAGGAA 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       4920 TAGAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCCCAGTAGGAA 4979

Qy       1981 ACAGTTCTTCTAAAGCCTATT 2001
              ||||||||||||||||||||
Db       4980 ACAGTTCTTCTAAAGCCTATT 5000


RESULT 2
AL583850
LOCUS       AL583850               165329 bp    DNA     linear   PRI 15-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-430G6 on chromosome 1, complete
            sequence.
ACCESSION   AL583850
VERSION     AL583850.5  GI:16973044
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165329)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 16, 2001 this sequence version replaced gi:15020514.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-430G6 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-430G6 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP11-430G6 is at 165329 in this
            sequence. The true right end of clone RP11-331H2 is at 2000 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .165329
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-430G6"
                     /clone_lib="RPCI-11.2"
BASE COUNT    51144 a  30897 c  31439 g  51849 t
ORIGIN

  Query Match             99.9%;  Score 1999.4;  DB 9;  Length 165329;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2000;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy          1 ATTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74681 ATTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAG 74740

Qy         61 AGAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74741 AGAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAG 74800

Qy        121 TCATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74801 TCATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGT 74860

Qy        181 CATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74861 CATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAA 74920

Qy        241 TATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74921 TATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGC 74980

Qy        301 AGGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74981 AGGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTA 75040

Qy        361 TAATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75041 TAATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAAT 75100

Qy        421 AGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75101 AGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTA 75160

Qy        481 CATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75161 CATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTC 75220

Qy        541 ACTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75221 ACTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTG 75280

Qy        601 CCCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75281 CCCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTAC 75340

Qy        661 CCACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75341 CCACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACC 75400

Qy        721 TTTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75401 TTTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCA 75460

Qy        781 AATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75461 AATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACT 75520

Qy        841 TCTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75521 TCTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTG 75580

Qy        901 ACCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75581 ACCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAA 75640

Qy        961 TGTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75641 TGTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAG 75700

Qy       1021 TGGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75701 TGGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATG 75760

Qy       1081 CTAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAAC 1140
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75761 CTAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAAC 75820

Qy       1141 TAGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAAT 1200
                ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db      75821 TAGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTAGAATTGGAAT 75880

Qy       1201 CCAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCA 1260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75881 CCAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCA 75940

Qy       1261 CCTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGC 1320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75941 CCTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGC 76000

Qy       1321 AGAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAA 1380
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76001 AGAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAA 76060

Qy       1381 ACAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGA 1440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76061 ACAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGA 76120

Qy       1441 TGCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCA 1500
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76121 TGCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCA 76180

Qy       1501 CAAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCC 1560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76181 CAAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCC 76240

Qy       1561 TGTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAA 1620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76241 TGTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAA 76300

Qy       1621 AATCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGG 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76301 AATCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGG 76360

Qy       1681 GAGGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGC 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76361 GAGGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGC 76420

Qy       1741 TGAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAG 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76421 TGAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAG 76480

Qy       1801 AACTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATC 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76481 AACTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATC 76540

Qy       1861 CTTGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACC 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76541 CTTGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACC 76600

Qy       1921 TAGAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCCCAGTAGGAA 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76601 TAGAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCCCAGTAGGAA 76660

Qy       1981 ACAGTTCTTCTAAAGCCTATT 2001
                |||||||||||||||||||
Db      76661 ACAGTTCTTCTAAAGCCTATT 76681


RESULT 3
AC031977/c
LOCUS       AC031977               191699 bp    DNA     linear   HTG 12-APR-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-288O18, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC031977
VERSION     AC031977.7  GI:13194952
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
            Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
            Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
            Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Mar 4, 2001 this sequence version replaced gi:9665085.
            -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 880
                 Center clone name: RP11-288O18
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513; 100% of reads
                 Sequencing Vector: plasmid; plasmid_accession; 0% of reads
                 Chemistry: Dye-primer; 1% of reads
                 Chemistry: Dye-terminator Big Dye; 99% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 190680 bases at least Q40
                 Consensus quality: 191287 bases at least Q30
                 Consensus quality: 191336 bases at least Q20
                 Insert size: 195548; agarose-fp
                 Insert size: 191499; sum-of-contigs
                 Quality coverage:  7.9x in Q20 bases; agarose-fp
                 Quality coverage:  8.1x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    12646: contig of 12646 bp in length
            *   12647    12746: gap of unknown length
            *   12747    94961: contig of 82215 bp in length
            *   94962    95061: gap of unknown length
            *   95062   191699: contig of 96638 bp in length.
FEATURES             Location/Qualifiers
     source          1. .191699
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-288O18"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .12646
                     /note="assembly_name:Contig5"
     misc_feature    12747. .94961
                     /note="assembly_name:Contig6
                     clone_end:T7"
     misc_feature    95062. .191699
                     /note="assembly_name:Contig7
                     clone_end:SP6"
BASE COUNT    57654 a  36385 c  36166 g  61293 t    201 others
ORIGIN
```

```
  Query Match             99.5%;  Score 1991.4;  DB 2;  Length 191699;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1995;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 ATTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9938 ATTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAG 9879

Qy         61 AGAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9878 AGAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAG 9819

Qy        121 TCATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9818 TCATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGT 9759

Qy        181 CATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9758 CATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAA 9699

Qy        241 TATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9698 TATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGC 9639

Qy        301 AGGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9638 AGGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTA 9579

Qy        361 TAATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9578 TAATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAAT 9519

Qy        421 AGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9518 AGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTA 9459

Qy        481 CATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9458 CATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTC 9399

Qy        541 ACTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9398 ACTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTG 9339

Qy        601 CCCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9338 CCCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTAC 9279

Qy        661 CCACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9278 CCACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACC 9219

Qy        721 TTTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9218 TTTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCA 9159

Qy        781 AATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9158 AATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACT 9099

Qy        841 TCTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9098 TCTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTG 9039

Qy        901 ACCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9038 ACCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAA 8979

Qy        961 TGTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8978 TGTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAG 8919
```

```
Qy       1021 TGGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8918 TGGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATG 8859

Qy       1081 CTAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAAC 1140
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       8858 CTAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCAGGGCACGTTACAAAAC 8799

Qy       1141 TAGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAAT 1200
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       8798 TAGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTAGAATTGGAAT 8739

Qy       1201 CCAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8738 CCAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCA 8679

Qy       1261 CCTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8678 CCTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGC 8619

Qy       1321 AGAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAA 1380
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       8618 AGAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTCCTTCTTGGGAA 8559

Qy       1381 ACAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8558 ACAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGA 8499

Qy       1441 TGCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8498 TGCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCA 8439

Qy       1501 CAAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8438 CAAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCC 8379

Qy       1561 TGTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8378 TGTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAA 8319

Qy       1621 AATCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGG 1680
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8318 ACTCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGG 8259

Qy       1681 GAGGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8258 GAGGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGC 8199

Qy       1741 TGAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAG 1800
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       8198 TGAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAATATCTGGTAG 8139

Qy       1801 AACTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATC 1860
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       8138 AACTAAAGAACGCAACTCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATC 8079

Qy       1861 CTTGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8078 CTTGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACC 8019

Qy       1921 TAGAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCCCAGTAGGAA 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8018 TAGAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCCCAGTAGGAA 7959

Qy       1981 ACAGTTCTTCTAAAGCCTATT 2001
              |||||||||||||||||||||
Db       7958 ACAGTTCTTCTAAAGCCTATT 7938
```

```
RESULT 4
AC111539/c
LOCUS       AC111539              234560 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-158A12, WORKING DRAFT SEQUENCE, 3
```

unordered pieces.
ACCESSION AC111539
VERSION AC111539.5 GI:30579258
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 234560)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 234560)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 234560)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On May 13, 2003 this sequence version replaced gi:25007559.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GMNF
Center clone name: CH230-158A12
-------------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 221222 bases at least Q40
Consensus quality: 224241 bases at least Q30
Consensus quality: 226292 bases at least Q20
Estimated insert size: 231347; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 231343: contig of 231343 bp in length
* 231344 231443: gap of unknown length
* 231444 232446: contig of 1003 bp in length
* 232447 232546: gap of unknown length
* 232547 234560: contig of 2014 bp in length.

FEATURES Location/Qualifiers
source 1. .234560
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-158A12"
misc_feature 1. .1537
/note="wgs_contig"
misc_feature 3731. .7474
/note="wgs_contig"
misc_feature 225356. .228724
/note="wgs_contig"
misc_feature 230080. .231343
/note="wgs_contig"

BASE COUNT 65515 a 49188 c 46995 g 66014 t 6848 others
ORIGIN

Query Match 11.6%; Score 231.2; DB 2; Length 234560;
Best Local Similarity 63.7%; Pred. No. 3.1e-43;
Matches 503; Conservative 0; Mismatches 238; Indels 49; Gaps 8;

```
Qy       1182 GAGGATCTATAATTGGAATCCAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGAC 1241
             ||||| | | ||||  |||| |  ||||||||   || | ||| ||||||  |   |||
Db       7000 GAGGAGCAAGAATTAAAATCTAAGTCTGCCTCCTGGTGAGGTTTAAGCACATGATATG-- 6943

Qy       1242 ACCATACTGTTTCTTTCCACCTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCT 1301
                 |||  |||| ||||||||||||||||||||||| |||||||| || |||||||
Db       6942 ------CTGCCTCTTCCCACCTGCACAATGCAAATGATCTCTTATGTAATTGCTGTTCTA 6889
```

```
Qy       1302 ATCCTGGGCTAAATGTTGCAGAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTT 1361
Db       6888 TATTGGGCTAAGCCCTTCCCAAGAAAGAGAACAATCTTTGGAATAAGATTGTTTTGGTTC 6829

Qy       1362 TTCTCCTACTTCTTGGGAAACAA--GGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTG 1419
Db       6828 ATCCTCCACTTCTTGGAAAAGAAAGGCTCTGTGTCCCCCAGCTAGTTAAGCATGTGACAG 6769

Qy       1420 TTCTTCCAGGGAAATCAGTGATGCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTC 1479
Db       6768 TTCTTCTAG-------AGTGATGCATCACCTGCTCTTGTCAAGTGTCAGGGTCAGAGTCC 6716

Qy       1480 CTGATTTATTGCATGTGCCCACAAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAG 1539
Db       6715 TTGGTTTATTGC----------------------AGAGCCAAACGTGTTTCCCAAAAG 6680

Qy       1540 TAAGACATACTGGG-AAGTCCCTGTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCAT 1598
Db       6679 CCATGGATACTGGGAAAGTCCCTATGCGTATTCCTGGCATACAGCACACTCTGGCCTGGT 6620

Qy       1599 ATCTTTGCTTTTTAGTCCTAAAAATCAATAA--CTGAACTCTCATTGATGTCTAGGCCAT 1656
Db       6619 CTGTTTGCTTGTTATCCTTAAAAACTAACAATCATGAACTCTCACTGATGTCAGGGCCAC 6560

Qy       1657 TGTAGTAAACAATAAAGAAGGAGGGAGGCTTCTGACAACTGAGAGGAAATTGTCATCTGA 1716
Db       6559 TGTAGGAGACAATAAAGAGAGTAGAGGTTTTTTCACAGCTTAGAGGAAATTGTAATCTGA 6500

Qy       1717 AGTGGTGCAAGCACAGCCTGGGGCTGAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATC 1776
Db       6499 GGTAGAGCAAATATAACCTAAGGCTGAGTTGTGGCTCACATCCTGCCCTGGAGAGGCTGC 6440

Qy       1777 AGTGCCCCATTTAACATCTGGTAGAACTAAAGAA---CGCAACGCCTGCCACAATGACTT 1833
Db       6439 GATCTCCCTTTTAGAATCTGATGAACTCAGAGAAATTCAAAAGACCTGCCCCAGTGCCTT 6380

Qy       1834 ATTTCCCTGCATTTGATACCGTCAATCCTTGAGAAATGTTTTCTTTTGTTCTCCCTGAGC 1893
Db       6379 ATTTCCTTGTGTTAAAGATCACAAAAATACTTGAGAAAGATTCTTT--TTCACCAGGAGA 6322

Qy       1894 AAAGGTTGGAAAAATTTGAAATTTACCTAGAGACCACACATAGTTCACATCCTGCTGTGT 1953
Db       6321 GAAATCTGGCGGATTTCAAATTATATCAGTGTACCACACACAGCTCCCAGCACGCTGCGT 6262

Qy       1954 GGCTGAATGT 1963
Db       6261 GACTTCATGT 6252
```

RESULT 5
AC125563/c

```
LOCUS       AC125563               275631 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-9B12, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC125563
VERSION     AC125563.4  GI:24817906
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 275631)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 275631)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 275631)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:23096552.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDBK
```

```
            Center clone name: CH230-9B12
           ---------- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 228152 bases at least Q40
            Consensus quality: 230310 bases at least Q30
            Consensus quality: 231872 bases at least Q20
            Estimated insert size: 235029; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
           ----------
           * NOTE: Estimated insert size may differ from sequence length
           *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 4 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *       1   192878: contig of 192878 bp in length
           *  192879   192978: gap of unknown length
           *  192979   270295: contig of 77317 bp in length
           *  270296   270395: gap of unknown length
           *  270396   272405: contig of 2010 bp in length
           *  272406   272505: gap of unknown length
           *  272506   275631: contig of 3126 bp in length.
FEATURES             Location/Qualifiers
     source          1. .275631
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-9B12"
     misc_feature    30410. .30884
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH339524"
     misc_feature    complement(56153. .56817)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH339528"
     misc_feature    57750. .59243
                     /note="wgs_end_extension
                     clone_end:Sp6"
BASE COUNT    68681 a  48371 c  48052 g  67696 t  42831 others
ORIGIN

  Query Match             11.6%;  Score 231.2;  DB 2;  Length 275631;
  Best Local Similarity   63.7%;  Pred. No. 3.1e-43;
  Matches  503;  Conservative    0;  Mismatches  238;  Indels   49;  Gaps    8;

Qy       1182 GAGGATCTATAATTGGAATCCAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGAC 1241
             ||||| | | ||||  |||| |  ||||||||    || | ||| |||||| |   |||
Db     147760 GAGGAGCAAGAATTAAAATCTAAGTCTGCCTCCTGGTGAGGTTTAAGCACATGATATG-- 147703

Qy       1242 ACCATACTGTTTCTTTCCACCTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCT 1301
                 |||  |||| ||||||||||||||||||||| |||||||| || |||||||
Db     147702 ------CTGCCTCTTCCCACCTGCACAATGCAAATGATCTCTTATGTAATTGCTGTTCTA 147649

Qy       1302 ATCCTGGGCTAAATGTTGCAGAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTT 1361
                ||    |    || |  | |||      |||||||| |||||  | ||||||  |
Db     147648 TATTGGGCTAAGCCCTTCCCAAGAAAGAGAACAATCTTTGGAATAAGATTGTTTTGGTTC 147589

Qy       1362 TTCTCCTACTTCTTGGGAAACAA--GGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTG 1419
                ||  | |||||||| ||| ||  | | |   |||||  |||| |||||  ||  | |
Db     147588 ATCCTCCACTTCTTGGAAAAGAAAGGCTCTGTGTCCCCCAGCTAGTTAAGCATGTGACAG 147529

Qy       1420 TTCTTCCAGGGAAATCAGTGATGCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTC 1479
             ||||||  ||        ||||||||||||||||||  | |||| |||||||||  |||| |
Db     147528 TTCTTCTAG-------AGTGATGCATCACCTGCTCTTGTCAAGTGTCAGGGTCAGAGTCC 147476

Qy       1480 CTGATTTATTGCATGTGCCCACAAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAG 1539
```

```
                ||  ||||||||                          |||     ||   ||||||||||
Db     147475 TTGGTTTATTGC-----------------------AGAGCCAAACGTGTTTCCCAAAAG 147440

Qy       1540 TAAGACATACTGGG-AAGTCCCTGTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCAT 1598
                |  |||||||| |||||||| |     |||||| |||||||  |||  |||    |
Db     147439 CCATGGATACTGGGAAAGTCCCTATGCGTATTCCTGGCATACAGCACACTCTGGCCTGGT 147380

Qy       1599 ATCTTTGCTTTTTAGTCCTAAAAATCAATAA--CTGAACTCTCATTGATGTCTAGGCCAT 1656
                | ||||||| |||  | ||||||  || ||   |||||||||| |||||||  |||||
Db     147379 CTGTTTGCTTGTTATCCTTAAAAACTAACAATCATGAACTCTCACTGATGTCAGGGCCAC 147320

Qy       1657 TGTAGTAAACAATAAAGAAGGAGGGAGGCTTCTGACAACTGAGAGGAAATTGTCATCTGA 1716
                ||||| | ||||||||||  |  |  |  || | ||| || |||||||||||| ||||||
Db     147319 TGTAGGAGACAATAAAGAGAGTAGAGGTTTTTTCACAGCTTAGAGGAAATTGTAATCTGA 147260

Qy       1717 AGTGGTGCAAGCACAGCCTGGGGCTGAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATC 1776
                || | ||||  | | |||  |||||||   ||||  ||||||||||  | |  |   |
Db     147259 GGTAGAGCAAATATAACCTAAGGCTGAGTTGTGGCTCACATCCTGCCCTGGAGAGGCTGC 147200

Qy       1777 AGTGCCCCATTTAACATCTGGTAGAACTAAAGAA---CGCAACGCCTGCCACAATGACTT 1833
                |  ||| ||||  ||||| |  |   | ||||   |  ||  |||||| || || |||
Db     147199 GATCTCCCTTTTAGAATCTGATGAACTCAGAGAAATTCAAAAGACCTGCCCCAGTGCCTT 147140

Qy       1834 ATTTCCCTGCATTTGATACCGTCAATCCTTGAGAAATGTTTTCTTTTGTTCTCCCTGAGC 1893
                |||||| ||  ||  | | |  ||     || |   ||||||  ||| ||  |||
Db     147139 ATTTCCTTGTGTTAAAGATCACAAAAATACTTGAGAAAGATTCTTT--TTCACCAGGAGA 147082

Qy       1894 AAAGGTTGGAAAAATTTGAAATTTACCTAGAGACCACACATAGTTCACATCCTGCTGTGT 1953
                ||  |||   | ||  || | ||      |||||||| || || || |  |||| ||
Db     147081 GAAATCTGGCGGATTTCAAATTATATCAGTGTACCACACACAGCTCCCAGCACGCTGCGT 147022

Qy       1954 GGCTGAATGT 1963
                | ||  ||||
Db     147021 GACTTCATGT 147012
```

```
RESULT 6
AC013439/c
LOCUS       AC013439              167891 bp    DNA     linear   PRI 07-NOV-2001
DEFINITION  Homo sapiens BAC clone RP11-270G18 from 2, complete sequence.
ACCESSION   AC013439
VERSION     AC013439.11  GI:13270751
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167891)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 167891)
  AUTHORS   Nguyen,C., Cotton,M., Hawkins,M. and Spalding,L.
  TITLE     The sequence of Homo sapiens BAC clone RP11-270G18
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 167891)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 167891)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 167891)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
```

COMMENT On Mar 10, 2001 this sequence version replaced gi:12280930.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_NH0270G18
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is AC068833; the clone sequenced to the right is RP11-88L20, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-270G18; actual end is at base position 167891 of RP11-270G18.

FEATURES Location/Qualifiers

source 1. .167891
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-270G18"
/clone_lib="RPCI-11"

repeat_region 1. .157
/rpt_family="L1"

repeat_region 155. .356
/rpt_family="L1"

repeat_region 357. .529
/rpt_family="L1"

repeat_region 530. .1024
/rpt_family="L1"

repeat_region 1053. .1237
/rpt_family="MIR"

repeat_region 1289. .1449
/rpt_family="MER1_type"

repeat_region 1965. .2014
/rpt_family="AT_rich"

repeat_region 2016. .2795
/rpt_family="L1"

misc_feature 2502. .2703
/note="similar to EST AI418310 (NID:g4264241) tg36d06.x1"

repeat_region 2796. .3345
/rpt_family="L2"

repeat_region 4244. .4566
/rpt_family="MER2_type"

misc_feature 5444. .5570
/note="similar to EST BE247703 (NID:g9099550)"

misc_feature 5464. .5570
/note="similar to EST AV727426 (NID:g10836847)"

misc_feature 5465. .5570
/note="similar to EST BE242721 (NID:g9094452)"

misc_feature 5466. .5570
/note="similar to EST AU124827 (NID:g10949543)"

misc_feature 5467. .5570
/note="similar to EST AA087204 (NID:g1630415) mk22h02.r1"

misc_feature 5470. .5570
/note="similar to EST BF693237 (NID:g11978645)"

misc_feature 5477. .5570
/note="similar to EST AU132620 (NID:g10993159)"

misc_feature 5491. .5570
/note="similar to EST AA334300 (NID:g1986543)"

misc_feature 5491. .5570
/note="similar to EST BG035852 (NID:g12430401)"

misc_feature 5497. .5570
/note="similar to EST BE891195 (NID:g10350282)"

misc_feature 5501. .5570
/note="similar to EST AL543591 (NID:g12876070)"

misc_feature 5514. .5570
/note="similar to EST BE783702 (NID:g10204900)"

repeat_region 6090. .6311
/rpt_family="MIR"

repeat_region 6432. .6737
/rpt_family="Alu"

repeat_region 7097. .7431
/rpt_family="L2"

repeat_region 7829. .7966
/rpt_family="L2"

repeat_region 7964. .8131
/rpt_family="L2"

repeat_region 9070. .9257
/rpt_family="MIR"

repeat_region 9293. .9490
/rpt_family="L2"

repeat_region 9496. .9800
/rpt_family="L1"

repeat_region 10495. .10599
/rpt_family="Tc2"

repeat_region 10646. .11164
/rpt_family="MER2_type"

repeat_region 11852. .12152
/rpt_family="MaLR"

repeat_region 12207. .12297
/rpt_family="Alu"

repeat_region 12299. .12334
/rpt_family="MaLR"

misc_feature 12365. .12500
/note="similar to EST BF158778 (NID:g11038867)"

misc_feature 12365. .12500
/note="similar to EST BG088353 (NID:g12570917)"

misc_feature 12371. .12500
/note="similar to EST AA087204 (NID:g1630415) mk22h02.r1"

misc_feature 12371. .12500
/note="similar to EST AL543591 (NID:g12876070)"

misc_feature 12371. .12500
/note="similar to EST AU124827 (NID:g10949543)"

misc_feature 12371. .12500
/note="similar to EST AU132620 (NID:g10993159)"

misc_feature 12371. .12500
/note="similar to EST AV727426 (NID:g10836847)"

misc_feature 12371. .12500
/note="similar to EST BE242721 (NID:g9094452)"

misc_feature 12371. .12500
/note="similar to EST BE247703 (NID:g9099550)"

misc_feature 12371. .12500
/note="similar to EST BE783702 (NID:g10204900)"

```
                /note="similar to  EST BE891195 (NID:g10350282)"
     misc_feature    12371. .12500
                /note="similar to  EST BF693237 (NID:g11978645)"
     misc_feature    12371. .12500
                /note="similar to  EST BG035852 (NID:g12430401)"
     misc_feature    12371. .12438
                /note="similar to  EST AA334300 (NID:g1986543)"
     misc_feature    12377. .12500
                /note="similar to  EST AW121877 (NID:g6097251)"
     misc_feature    12397. .12500
                /note="similar to  EST BF334128 (NID:g11304876)"
     misc_feature    12409. .12500
                /note="similar to  EST AI556441 (NID:g4488804)"
     repeat_region   12730. .12862
                /rpt_family="MER103"
     misc_feature    12793. .12826
                /note="similar to  EST BE242721 (NID:g9094452)"
     misc_feature    13237. .13277
                /note="similar to  EST BE242721 (NID:g9094452)"

  Query Match             9.8%;  Score 196.4;  DB 9;  Length 167891;
  Best Local Similarity   74.0%;  Pred. No. 3.8e-35;
  Matches  305;  Conservative    0;  Mismatches   96;  Indels   11;  Gaps    4;

Qy        169 TTTATTTACAGTCATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTAT 228
              ||| |||||||||||| ||||||||  |   ||| ||||  || ||| |||| || |
Db     162394 TTTGATTACAGTCATGTACCACATAATGACCCTTCGGTCAAAGACAGACCACATATATGA 162335

Qy        229 CTGGCCCCATAATATTATA------AGCTGAGAAATTTCTATTAACTAGTGATATCGCAG 282
                |   |  ||  |   ||       |||||||||||| |||| | |||||||| ||
Db     162334 TGGTGGCTCTATAAGATTACTATGGAGCTGAGAAATTCCTATCATCTAGTGATGTCATTA 162275

Qy        283 CCATC---ATAAGTGTAATGCAGGACATTACCTTTTCTATGTTTAGATATG-TTAGATAC 338
              |       | ||   || ||||   ||||| |||||||||||||||||||| ||| || |
Db     162274 GCTGTTGTAAAATCATAGTGCAATGCATTATCTTTTCTATGTTTAGATATGTTTAAATTC 162215

Qy        339 ACAAATATATTTCATTGTGTTATAATTTCCTACAGTATTCAGTACAGTAACATGCTGTAC 398
              ||||||| |     ||||| |  | | |||||||||| ||||||||||||||||||||
Db     162214 ACAAATACTTACTGCTGTGTAAGTACTGCCTACAGTATCCAGTACAGTAACATGCTGTAC 162155

Qy        399 AGGTTTGTAACCTAGGAGTAATAGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATA 458
              ||||||||| |||||||| || |||||||||  || ||| ||||||||||||||||||
Db     162154 AGGTTTGTAGCCTAGGAGCAACAGGCTATACTCTATAGCCTAGGTGTGTAGTAGGCTAT- 162096

Qy        459 ACCATCTAGGTTTGTGTAAGTACATTCTATGATATTCCCACAATGATGAAATCACCTAAC 518
              || ||| |||||||| |||||||| |||||||| ||  |||||||  ||||  |||||
Db     162095 ACAATCCAGGTTTGTCTAAGTACACTCTATGATGTTTGCACAATGACAAAATTGCCTAAT 162036

Qy        519 TACACATTTCTCAGAATGTTTCACTGTTGTGAAGTGACCCATGACTATATTT 570
                |   |||||||||||||| || ||| ||| ||||||   |||||||| ||
Db     162035 GATGAATTTCTCAGAATGTATCCCTGCTGTTAAGTGATGTATGACTATACTT 161984


RESULT 7
HS55I10/c
LOCUS       HS55I10                 90429 bp    DNA     linear   PRI 04-MAR-2003
DEFINITION  Human DNA sequence from clone CTA-55I10 on chromosome 1q32.2-41,
            complete sequence.
ACCESSION   AL035408
VERSION     AL035408.5  GI:4775616
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 90429)
  AUTHORS   Bird,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 11, 1999 this sequence version replaced gi:4584773.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            CTA-55I10 is from the human BAC library described in U-J. Kim et
            al. (1996) Genomics 34, 213-218.
            VECTOR: pBAC108L.
FEATURES            Location/Qualifiers
     source         1. .90429
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="1"
                    /map="q32.2-41"
                    /clone="CTA-55I10"
                    /clone_lib="CIT978SK-A1"
BASE COUNT    26268 a  20370 c  18924 g  24867 t
ORIGIN

  Query Match             9.6%;  Score 192;  DB 9;  Length 90429;
  Best Local Similarity   74.0%;  Pred. No. 4.3e-34;
  Matches  311;  Conservative    0;  Mismatches  100;  Indels    9;  Gaps    5;

Qy        148 CTTATCACAATTTATCATTATTTTATTTACAGTCATGTGCCACATAACAATGTTTCAGTC 207
              |||||  |   | |     ||| |  |||||||     ||||||||  || |||||||||
Db      79639 CTTATGCCCTCTACTTTCATTTTCAACTACAGTCACACACCACATAATGATATTTCAGTC 79580

Qy        208 AGGGATAGAACACAAATGTATC--TGGCCCCATAATATTATA----AGCTGAGAAATTTC 261
              |  || | | |||| ||| | |  ||| ||||||| ||||||    |||||| ||||| |
Db      79579 ATTGACAAACCACATATGCAACATTGGTCCCATAAGATTATAAGGGAGCTGAAAAATTCC 79520

Qy        262 TATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCAGGACATTACCTTTTCTATGT 321
              | ||  ||||||| || ||   ||| | |   |     ||  |||| || ||||| | ||
Db      79519 TGTTGCCTAGTGACATAGCTTTCAT-AACATCATGGCTCAACACATAACTTTTTCCAAGT 79461

Qy        322 TTAGATA-TGTTAGATACACAAATATATTTCATTGTGTTATAATTTCCTACAGTATTCAG 380
              |||||||   ||| |||||||||| |  |||||||||| |  |  || ||||||||||
Db      79460 TTAGATACATTTAAATACACAAATACTTACCATTGTGTTACAGATGTCTGCAGTATTCAG 79401

Qy        381 TACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAGGCTATACCATACAGCTTA 440
              |||||||||||| |||||||||||| ||| |||||||||||| |||||||| ||||||
Db      79400 TACAGTAACATGCAGTACAGGTTTGTAGCCTGGGAGTAATAGGCCATACCATATAGCTTA 79341

Qy        441 GGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACATTCTATGATATTCCCACA 500
              || |||||||||||||| |||||| |||||||    | |||    |||||| | ||  ||||
Db      79340 GGCGTGTAGTAGGCTAT-ACCATCAAGGTTTGCAGATGTATGCTCTATGGTGTTTGCACA 79282

Qy        501 ATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCACTGTTGTGAAGTGACCCAT 560
              ||||  |||||||||| |  |||||||||||| | |  ||||||| ||||||| |||
```

```
Db      79281 ATGACAAAATCACCTAAAGATGCATTTCTCAGAATATGTTCCTGTTGTTAAGTGACACAT 79222


RESULT 8
AC120598_2
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AC120598 Accession AC120598
     Fragment Name      Begin      End
     AC120598_0             1    110000
     AC120598_1        100001    210000
     AC120598_2        200001    310000
     AC120598_3        300001    367307
Continuation (3 of 4) of AC120598 from base 200001 (AC120598 Homo sapiens, *** SEQUENCIN

  Query Match             9.6%;  Score 192;  DB 2;  Length 110000;
  Best Local Similarity   70.7%;  Pred. No. 4.2e-34;
  Matches  328;  Conservative    0;  Mismatches  125;  Indels   11;  Gaps    5;

Qy        117 TAAGTCATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTA 176
              ||| ||| |  | | |||   |  |     |  |||   ||   |      |  | | ||
Db      94957 TAACTCAGCATTATCTCAACATTGTTTTTTATTTACTGGAAAAAAGTCCCAGGTACTATG 95016

Qy        177 CAGTCATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTA--TCTGGCC 234
              ||||||| |   |||||  ||| |||||||||  |  || || |  | ||    ||| |
Db      95017 TAGTCATGAGGTGCATAATGATGCTTCAGTCAGCAAAGGACCATATTTATACTAGTGGTC 95076

Qy        235 CCATAATATTATA----AGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATA 290
              |||| | | ||||    |||||| ||||| || | | |||||||  ||  ||||||||||
Db      95077 CCATGAGAATATAATGGAGCTGAAAAATTCCTGTCACCTAGTGAGGTCTTAGCCATCATA 95136

Qy        291 AG---TGTAATGCAGGACATTACCTTTTCTATGTTTAGATAT-GTTAGATACACAAATAT 346
              |    |||  ||||   ||||| |||||||||||||||||  ||||||||||||||||
Db      95137 ATGACTGTGGTGCAATGCATTATCTTTTCTATGTTTAGATATATTTAGATACACAAATAC 95196

Qy        347 ATTTCATTGTGTTATAATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGT 406
              |  |||||||||| | || ||||||| | |||||||||||||||||||| | ||||||
Db      95197 TTACCATTGTGTTACAGTTGCCTACAGCACTCAGTACAGTAACATGCTGTATAAGTTTGT 95256

Qy        407 AACCTAGGAGTAATAGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTA 466
              | | ||||| |||||||||| |||   ||| |||||||||| | | |   ||||||||
Db      95257 AGCTTAGGATTAATAGGCTATGCCACGTAGCCTAGGTGTGTAGCAAG-TGGTACCATCTA 95315

Qy        467 GGTTTGTGTAAGTACATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATT 526
              |||||||||||||||| || ||||| ||   ||||||| ||||||  |||     |||
Db      95316 GGTTTGTGTAAGTACACTCCATGATGTTTGTACAATGACAAAATCAGTTAATGGTGCATC 95375

Qy        527 TCTCAGAATGTTTCACTGTTGTGAAGTGACCCATGACTATATTT 570
              ||||||||  | |  ||||||| || |||| ||| ||||||  ||
Db      95376 TCTCAGAACATATTCCTGTTGTTAAATGACACATAACTATACTT 95419


RESULT 9
AC069222
LOCUS       AC069222               117000 bp    DNA     linear   PRI 29-MAR-2002
DEFINITION  Homo sapiens 3 BAC RP11-383I23 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC069222
VERSION     AC069222.23  GI:19033388
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 117000)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 117000)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 117000)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 117000)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 117000)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 1, 2002 this sequence version replaced gi:18958589.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
```

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT.

```
FEATURES             Location/Qualifiers
     source          1. .117000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-383I23"
     misc_feature    complement(1. .1988)
                     /note="overlaps bases 1. .1988 of clone AC024938"
                     /function="clone overlap"
     repeat_region   2059. .2084
                     /rpt_family="AT_rich"
     repeat_region   2505. .2662
                     /rpt_family="MIR"
     repeat_region   complement(2805. .2881)
                     /rpt_family="L2"
     repeat_region   complement(2936. .3038)
                     /rpt_family="L2"
     repeat_region   complement(3024. .3183)
                     /rpt_family="L2"
     repeat_region   3300. .3464
                     /rpt_family="MIR"
     repeat_region   4436. .4779
                     /rpt_family="MER1B"
     repeat_region   5133. .5358
                     /rpt_family="MER20"
     repeat_region   complement(5611. .5770)
                     /rpt_family="L1MA8"
     repeat_region   complement(5809. .6018)
                     /rpt_family="MIR"
     repeat_region   8572. .8611
                     /rpt_family="L2"
     repeat_region   9064. .9344
                     /rpt_family="Charlie1b"
     repeat_region   complement(9345. .9645)
                     /rpt_family="AluY"
     repeat_region   9646. .9790
                     /rpt_family="Charlie1b"
     repeat_region   complement(10718. .10878)
                     /rpt_family="L2"
     repeat_region   11415. .11535
                     /rpt_family="Charlie2"
     repeat_region   11689. .11731
                     /rpt_family="(CAGAG)n"
     repeat_region   11973. .12104
                     /rpt_family="Charlie2"
     repeat_region   12285. .12440
                     /rpt_family="MIR"
     repeat_region   12802. .12940
                     /rpt_family="MER5A"
     repeat_region   13114. .13290
                     /rpt_family="MER20"
     repeat_region   15088. .15367
                     /rpt_family="AluSc"
     repeat_region   15368. .15395
                     /rpt_family="(TAAAA)n"
     repeat_region   complement(16209. .16419)
                     /rpt_family="MIR"
     STS             19163. .19302
                     /standard_name="72547"
     STS             19297. .19426
                     /standard_name="171033"
     repeat_region   20164. .20228
                     /rpt_family="L2"
     STS             20770. .20907
                     /standard_name="43872"
     repeat_region   complement(22803. .23127)
                     /rpt_family="THE1C"
     repeat_region   complement(23481. .23607)
                     /rpt_family="MER5A"
     repeat_region   23517. .23626
                     /rpt_family="MER5A"
     repeat_region   23631. .23770
                     /rpt_family="MER47B"
     repeat_region   complement(23771. .24106)
                     /rpt_family="MER1B"
     repeat_region   24107. .24307
                     /rpt_family="MER47B"
     repeat_region   complement(25342. .26197)
                     /rpt_family="L1PB1"
     repeat_region   27024. .27060
                     /rpt_family="MIR"
     repeat_region   27309. .27608
```

```
Query Match             9.6%;  Score 192;  DB 9;  Length 117000;
Best Local Similarity   70.7%;  Pred. No. 4.2e-34;
Matches  328;  Conservative    0;  Mismatches  125;  Indels   11;  Gaps    5;

Qy        117 TAAGTCATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTA 176
              ||| ||| |  | | |||   |  |     |  |||    ||   |     |  |  | ||
Db      82983 TAACTCAGCATTATCTCAACATTGTTTTTTATTTACTGGAAAAAAGTCCCAGGTACTATG 83042

Qy        177 CAGTCATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTA--TCTGGCC 234
              ||||||| |   ||||| ||| |||||||||  |  || || |  | ||    ||| |
Db      83043 TAGTCATGAGGTGCATAATGATGCTTCAGTCAGCAAAGGACCATATTTATACTAGTGGTC 83102

Qy        235 CCATAATATTATA----AGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATA 290
              |||| | | ||||    |||||| ||||| || | | |||||||  ||  ||||||||||
Db      83103 CCATGAGAATATAATGGAGCTGAAAAATTCCTGTCACCTAGTGAGGTCTTAGCCATCATA 83162

Qy        291 AG---TGTAATGCAGGACATTACCTTTTCTATGTTTAGATAT-GTTAGATACACAAATAT 346
              |    |||  ||||   ||||| |||||||||||||||||||  ||||||||||||||
Db      83163 ATGACTGTGGTGCAATGCATTATCTTTTCTATGTTTAGATATATTTAGATACACAAATAC 83222

Qy        347 ATTTCATTGTGTTATAATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGT 406
              |  |||||||||| | || ||||||| | ||||||||||||||||||||||| | ||||||
Db      83223 TTACCATTGTGTTACAGTTGCCTACAGCACTCAGTACAGTAACATGCTGTATAAGTTTGT 83282

Qy        407 AACCTAGGAGTAATAGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTA 466
              | | ||||| |||||||||| ||||  |||  |||||||||| | | |   ||||||||
Db      83283 AGCTTAGGATTAATAGGCTATGCCATGTAGCCAAGGTGTGTAGCAAG-TGGTACCATCTA 83341

Qy        467 GGTTTGTGTAAGTACATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATT 526
              |||||||||||||||| || ||||| ||   |||||||  ||||||  |||      |||
Db      83342 GGTTTGTGTAAGTACACTCCATGATGTTTGTACAATGACAAAATCAGTTAATGGTGCATC 83401

Qy        527 TCTCAGAATGTTTCACTGTTGTGAAGTGACCCATGACTATATTT 570
              |||||||| | |  ||||||| || |||| ||| |||||| ||
Db      83402 TCTCAGAACATATTCCTGTTGTTAAATGACACATAACTATACTT 83445
```

RESULT 10

```
AC027182/c
LOCUS       AC027182              140952 bp    DNA     linear   HTG 28-MAR-2000
DEFINITION  Homo sapiens chromosome 1 clone RP11-154I19 map 1, WORKING DRAFT
            SEQUENCE, 39 unordered pieces.
ACCESSION   AC027182
VERSION     AC027182.1  GI:7331504
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 140952)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 1, clone RP11-154I19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 140952)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7203
                Center clone name: 154_I_19
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 123230 bases at least Q40
                Consensus quality: 131969 bases at least Q30
                Consensus quality: 134958 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 137152; sum-of-contigs
                Quality coverage: 2.6 in Q20 bases; agarose-fp
                Quality coverage: 3.0 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 39 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1256: contig of 1256 bp in length
            *     1257     1356: gap of 100 bp
            *     1357     2453: contig of 1097 bp in length
            *     2454     2553: gap of 100 bp
            *     2554     3853: contig of 1300 bp in length
            *     3854     3953: gap of 100 bp
            *     3954     5402: contig of 1449 bp in length
            *     5403     5502: gap of 100 bp
            *     5503     7076: contig of 1574 bp in length
            *     7077     7176: gap of 100 bp
            *     7177     8451: contig of 1275 bp in length
            *     8452     8551: gap of 100 bp
            *     8552     9933: contig of 1382 bp in length
            *     9934    10033: gap of 100 bp
            *    10034    11519: contig of 1486 bp in length
            *    11520    11619: gap of 100 bp
            *    11620    14442: contig of 2823 bp in length
            *    14443    14542: gap of 100 bp
            *    14543    15852: contig of 1310 bp in length
            *    15853    15952: gap of 100 bp
            *    15953    18160: contig of 2208 bp in length
            *    18161    18260: gap of 100 bp
            *    18261    20873: contig of 2613 bp in length
            *    20874    20973: gap of 100 bp
            *    20974    23073: contig of 2100 bp in length
            *    23074    23173: gap of 100 bp
            *    23174    25049: contig of 1876 bp in length
            *    25050    25149: gap of 100 bp
            *    25150    27256: contig of 2107 bp in length
            *    27257    27356: gap of 100 bp
            *    27357    29855: contig of 2499 bp in length
            *    29856    29955: gap of 100 bp
            *    29956    32443: contig of 2488 bp in length
            *    32444    32543: gap of 100 bp
            *    32544    36328: contig of 3785 bp in length
            *    36329    36428: gap of 100 bp
            *    36429    39039: contig of 2611 bp in length
            *    39040    39139: gap of 100 bp
            *    39140    42587: contig of 3448 bp in length
            *    42588    42687: gap of 100 bp
            *    42688    45620: contig of 2933 bp in length
            *    45621    45720: gap of 100 bp
            *    45721    48337: contig of 2617 bp in length
            *    48338    48437: gap of 100 bp
            *    48438    51930: contig of 3493 bp in length
            *    51931    52030: gap of 100 bp
            *    52031    55883: contig of 3853 bp in length
            *    55884    55983: gap of 100 bp
            *    55984    58926: contig of 2943 bp in length
            *    58927    59026: gap of 100 bp
            *    59027    63146: contig of 4120 bp in length
            *    63147    63246: gap of 100 bp
            *    63247    66639: contig of 3393 bp in length
            *    66640    66739: gap of 100 bp
            *    66740    70463: contig of 3724 bp in length
            *    70464    70563: gap of 100 bp
            *    70564    74363: contig of 3800 bp in length
            *    74364    74463: gap of 100 bp
            *    74464    79662: contig of 5199 bp in length
            *    79663    79762: gap of 100 bp
            *    79763    85047: contig of 5285 bp in length
            *    85048    85147: gap of 100 bp
            *    85148    89914: contig of 4767 bp in length
            *    89915    90014: gap of 100 bp
            *    90015    94457: contig of 4443 bp in length
            *    94458    94557: gap of 100 bp
            *    94558    99499: contig of 4942 bp in length
            *    99500    99599: gap of 100 bp
            *    99600   104279: contig of 4680 bp in length
            *   104280   104379: gap of 100 bp
            *   104380   110964: contig of 6585 bp in length
            *   110965   111064: gap of 100 bp
            *   111065   118181: contig of 7117 bp in length
            *   118182   118281: gap of 100 bp
            *   118282   128265: contig of 9984 bp in length
```

```
          *   128266   128365: gap of 100 bp
          *   128366   140952: contig of 12587 bp in length.
FEATURES             Location/Qualifiers
     source          1. .140952
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1"
                     /clone="RP11-154I19"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1256
                     /note="assembly_fragment"
     misc_feature    1357. .2453
                     /note="assembly_fragment"
     misc_feature    2554. .3853
                     /note="assembly_fragment"
     misc_feature    3954. .5402
                     /note="assembly_fragment"
     misc_feature    5503. .7076
                     /note="assembly_fragment"
     misc_feature    7177. .8451
                     /note="assembly_fragment"
     misc_feature    8552. .9933
                     /note="assembly_fragment"
     misc_feature    10034. .11519
                     /note="assembly_fragment"
     misc_feature    11620. .14442
                     /note="assembly_fragment"
     misc_feature    14543. .15852
                     /note="assembly_fragment"
     misc_feature    15953. .18160
                     /note="assembly_fragment"
     misc_feature    18261. .20873
                     /note="assembly_fragment"
     misc_feature    20974. .23073
                     /note="assembly_fragment"
     misc_feature    23174. .25049
                     /note="assembly_fragment"
     misc_feature    25150. .27256
                     /note="assembly_fragment"
     misc_feature    27357. .29855
                     /note="assembly_fragment"
     misc_feature    29956. .32443
                     /note="assembly_fragment"
     misc_feature    32544. .36328
                     /note="assembly_fragment"
     misc_feature    36429. .39039
                     /note="assembly_fragment"
     misc_feature    39140. .42587
                     /note="assembly_fragment"
     misc_feature    42688. .45620
                     /note="assembly_fragment"
     misc_feature    45721. .48337
                     /note="assembly_fragment"

  Query Match             9.6%;  Score 192;  DB 2;  Length 140952;
  Best Local Similarity   73.0%;  Pred. No. 4.1e-34;
  Matches  305;  Conservative    0;  Mismatches  100;  Indels   13;  Gaps    4;

Qy       170 TTATTTACAGTCATGTGCCACATAACAATGTTTCAGTCAGGGATAG--AACACAAATGTA 227
            ||   |||||||||||||  |||||||   |||||||||  |||  |    | ||| |||
Db     10694 TTTAATACAGTCATGTGCTGCATAACACCATTTCAGTCAATGATGGGCTATACATATGAC 10635

Qy       228 TCTGGCCCCATAATATTATAAG----CTGAGAAATTTCTATTAACTAGTGATATCGCAGC 283
            | ||| | ||||| |||||||     |||| |||||  ||    |||||| || | |||
Db     10634 TGTGGTCTCATAAAATTATAATGGAACTGAAAAATTCTTAAGGCCTAGTGACATTGTAGC 10575

Qy       284 CATCATAAG-TGTAATGCAGGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAA 342
            |||  ||| |   | ||   |||||||| ||| ||||||||||  |||| |||||||
Db     10574 CATGGTAACCTCCTTTTCATAGCATTACCTGTTCGATGTTTAGATAGTTTAGGTACACAA 10515

Qy       343 ATATATTTCATTGTGTTATAATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGT 402
            ||| |  ||| |||||| || | ||||||| | ||||||| ||||||||||||||||||
Db     10514 GTATTTACCATAGTGTTACAACTGCCTACAGAACTCAGTACTGTAACATGCTGTACAGGT 10455

Qy       403 TTGTAACCTAGGAGTAATAGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTAT----- 457
            | | | ||||  || ||||||| | ||||||   | ||||||| |||||||  ||
Db     10454 TGGCACCCTACCAGCAATAGGCCACACCATATGCCCTAGGTGTATAGTAGGGCATCTAGT 10395

Qy       458 -AACCATCTAGGTTTGTGTAAGTACATTCTATGATATTCCCACAATGATGAAATCACCTA 516
             |||||||| |||||||||| |||||||||||| ||  |  |||||  ||||||||||
Db     10394 ACACCATCTAGATTTGTGTAAGGACATTCTATGATGTTAACGTAATGACAAAATCACCTA 10335

Qy       517 ACTACACATTTCTCAGAATGTTTCACTGTTGTGAAGTGACCCATGACTATATTTTCCT 574
             |  |||| ||||||||||  | |  |  |||| ||||||| | ||||||||  | | |
Db     10334 ATGACACGTTTCTCAGAACATATTCCCATTGTTAAGTGACACCTGACTATACATCCAT 10277


RESULT 11
AC022883
LOCUS       AC022883              166992 bp    DNA     linear   HTG 16-MAR-2000
DEFINITION  Homo sapiens clone RP11-447O17, WORKING DRAFT SEQUENCE, 26
            unordered pieces.
ACCESSION   AC022883
VERSION     AC022883.3  GI:7249283
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166992)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-447O17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 166992)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 16, 2000 this sequence version replaced gi:6978240.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6218
                Center clone name: 447_O_17
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 152238 bases at least Q40
                Consensus quality: 159904 bases at least Q30
                Consensus quality: 162428 bases at least Q20
                Insert size: 170000; agarose-fp
```

```
            Insert size: 164492; sum-of-contigs
            Quality coverage: 3.5 in Q20 bases; agarose-fp
            Quality coverage: 3.6 in Q20 bases; sum-of-contigs
          ----------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 26 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1279: contig of 1279 bp in length
          *     1280     1379: gap of 100 bp
          *     1380     3074: contig of 1695 bp in length
          *     3075     3174: gap of 100 bp
          *     3175     4899: contig of 1725 bp in length
          *     4900     4999: gap of 100 bp
          *     5000     7406: contig of 2407 bp in length
          *     7407     7506: gap of 100 bp
          *     7507     9444: contig of 1938 bp in length
          *     9445     9544: gap of 100 bp
          *     9545    11702: contig of 2158 bp in length
          *    11703    11802: gap of 100 bp
          *    11803    14743: contig of 2941 bp in length
          *    14744    14843: gap of 100 bp
          *    14844    18243: contig of 3400 bp in length
          *    18244    18343: gap of 100 bp
          *    18344    19862: contig of 1519 bp in length
          *    19863    19962: gap of 100 bp
          *    19963    23231: contig of 3269 bp in length
          *    23232    23331: gap of 100 bp
          *    23332    26785: contig of 3454 bp in length
          *    26786    26885: gap of 100 bp
          *    26886    30112: contig of 3227 bp in length
          *    30113    30212: gap of 100 bp
          *    30213    33206: contig of 2994 bp in length
          *    33207    33306: gap of 100 bp
          *    33307    37673: contig of 4367 bp in length
          *    37674    37773: gap of 100 bp
          *    37774    41451: contig of 3678 bp in length
          *    41452    41551: gap of 100 bp
          *    41552    46864: contig of 5313 bp in length
          *    46865    46964: gap of 100 bp
          *    46965    52870: contig of 5906 bp in length
          *    52871    52970: gap of 100 bp
          *    52971    61002: contig of 8032 bp in length
          *    61003    61102: gap of 100 bp
          *    61103    70187: contig of 9085 bp in length
          *    70188    70287: gap of 100 bp
          *    70288    78232: contig of 7945 bp in length
          *    78233    78332: gap of 100 bp
          *    78333    88724: contig of 10392 bp in length
          *    88725    88824: gap of 100 bp
          *    88825   100892: contig of 12068 bp in length
          *   100893   100992: gap of 100 bp
          *   100993   114691: contig of 13699 bp in length
          *   114692   114791: gap of 100 bp
          *   114792   131880: contig of 17089 bp in length
          *   131881   131980: gap of 100 bp
          *   131981   147366: contig of 15386 bp in length
          *   147367   147466: gap of 100 bp
          *   147467   166992: contig of 19526 bp in length.
FEATURES             Location/Qualifiers
     source          1. .166992
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-447O17"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1279
                     /note="assembly_fragment"
     misc_feature    1380. .3074
                     /note="assembly_fragment"
     misc_feature    3175. .4899
                     /note="assembly_fragment"
     misc_feature    5000. .7406
                     /note="assembly_fragment"
     misc_feature    7507. .9444
                     /note="assembly_fragment"
     misc_feature    9545. .11702
                     /note="assembly_fragment"
     misc_feature    11803. .14743
                     /note="assembly_fragment"
     misc_feature    14844. .18243
                     /note="assembly_fragment"
     misc_feature    18344. .19862
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    19963. .23231
                     /note="assembly_fragment"
     misc_feature    23332. .26785
                     /note="assembly_fragment"
     misc_feature    26886. .30112
                     /note="assembly_fragment"
     misc_feature    30213. .33206
                     /note="assembly_fragment"
     misc_feature    33307. .37673
                     /note="assembly_fragment"
     misc_feature    37774. .41451
                     /note="assembly_fragment"
     misc_feature    41552. .46864
                     /note="assembly_fragment"
     misc_feature    46965. .52870
                     /note="assembly_fragment"
     misc_feature    52971. .61002
                     /note="assembly_fragment"
     misc_feature    61103. .70187
                     /note="assembly_fragment"
     misc_feature    70288. .78232
                     /note="assembly_fragment"
     misc_feature    78333. .88724
                     /note="assembly_fragment"
     misc_feature    88825. .100892
                     /note="assembly_fragment"
     misc_feature    100993. .114691
                     /note="assembly_fragment"
     misc_feature    114792. .131880
                     /note="assembly_fragment"
     misc_feature    131981. .147366
                     /note="assembly_fragment"
     misc_feature    147467. .166992
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
BASE COUNT    49197 a  32045 c  32526 g  50717 t   2507 others
ORIGIN

  Query Match             9.6%;  Score 192;  DB 2;  Length 166992;
  Best Local Similarity   70.7%;  Pred. No. 4e-34;
  Matches  328;  Conservative    0;  Mismatches  125;  Indels   11;  Gaps    5;

Qy        117 TAAGTCATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTA 176
              ||| ||| | | | |||   | | |  |  |||   ||  |   |  | | | |
Db        578 TAACTCAGCATTATCTCAACATTGTTTTTTATTTACTGGAAAAAAGTCCCAGGTACTATG 637

Qy        177 CAGTCATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTA--TCTGGCC 234
              ||||||| |   |||||  ||| |||||||||  |  || || |  | ||   ||| |
Db        638 TAGTCATGAGGTGCATAATGATGCTTCAGTCAGCAAAGGACCATATTTATACTAGTGGTC 697

Qy        235 CCATAATATTATA----AGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATA 290
              |||| | | ||||    |||||| ||||| || | | |||||||  ||  ||||||||||
Db        698 CCATGAGAATATAATGGAGCTGAAAAATTCCTGTCACCTAGTGAGGTCTTAGCCATCATA 757

Qy        291 AG---TGTAATGCAGGACATTACCTTTTCTATGTTTAGATAT-GTTAGATACACAAATAT 346
```

```
                |    |||  ||||   ||||| ||||||||||||||||||  ||||||||||||||||
Db        758 ATGACTGTGGTGCAATGCATTATCTTTTCTATGTTTAGATATATTTAGATACACAAATAC 817

Qy        347 ATTTCATTGTGTTATAATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGT 406
              |  |||||||||| | || ||||||| | ||||||||||||||||||| | ||||||
Db        818 TTACCATTGTGTTACAGTTGCCTACAGCACTCAGTACAGTAACATGCTGTATAAGTTTGT 877

Qy        407 AACCTAGGAGTAATAGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTA 466
              | | ||||| |||||||||| ||||  |||  |||||||||| | | |   ||||||||
Db        878 AGCTTAGGATTAATAGGCTATGCCATGTAGCCAAGGTGTGTAGCAAG-TGGTACCATCTA 936

Qy        467 GGTTTGTGTAAGTACATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATT 526
              |||||||||||||||| || ||||| ||   |||||||  ||||||  |||     |||
Db        937 GGTTTGTGTAAGTACACTCCATGATGTTTGTACAATGACAAAATCAGTTAATGGTGCATC 996

Qy        527 TCTCAGAATGTTTCACTGTTGTGAAGTGACCCATGACTATATTT 570
              |||||||| | |  ||||||| || |||| ||| ||||||  ||
Db        997 TCTCAGAACATATTCCTGTTGTTAAATGACACATAACTATACTT 1040
```

```
RESULT 12
AL133383/c
LOCUS       AL133383              204158 bp    DNA     linear   PRI 04-APR-2001
DEFINITION  Human DNA sequence from clone GS1-204I12 on chromosome 1. Contains
            ESTs, STSs and GSSs. Contains a novel gene, complete sequence.
ACCESSION   AL133383
VERSION     AL133383.10  GI:10119658
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 204158)
  AUTHORS   Hall,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: cloneregquest@sanger.ac.uk
COMMENT     On Sep 12, 2000 this sequence version replaced gi:10086005.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep GS1-204I12 is
            from the library Genome_Systems_ReleaseI VECTOR: pBeloBAC11
            This sequence is the entire insert of clone GS1-204I12 This
            sequence has been finished according to sequence map criteria as
            follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            This sequence was generated by the Sanger Centre from part of a
            human chromosome 1 bacterial clone contig constructed by John
            Carpten et al, NHGRI, NIH. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1.
FEATURES             Location/Qualifiers
     source          1. .204158
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="GS1-204I12"
                     /clone_lib="Genome_Systems_ReleaseI"
     misc_feature    7. .442
                     /note="match: GSS: Em:AQ761971"
     repeat_region   245. .300
                     /note="14 copies 4 mer tata 73% conserved"
     repeat_region   251. .298
                     /note="6 copies 8 mer tatacata 85% conserved"
     repeat_region   256. .301
                     /note="23 copies 2 mer at 80% conserved"
     repeat_region   313. .989
                     /note="L1ME3A repeat: matches 5406. .6152 of consensus"
     misc_feature    1047. .1513
                     /note="match: GSS: Em:AQ819686"
     repeat_region   1908. .2053
                     /note="FAM repeat: matches 26. .172 of consensus"
     repeat_region   2276. .2574
                     /note="AluSx repeat: matches 9. .303 of consensus"
     repeat_region   2720. .2870
                     /note="L2 repeat: matches 2558. .2711 of consensus"
     repeat_region   4330. .4437
                     /note="MER94 repeat: matches 10. .121 of consensus"
     repeat_region   4758. .4793
                     /note="18 copies 2 mer tt 91% conserved"
     repeat_region   5066. .5201
                     /note="MIR repeat: matches 7. .140 of consensus"
     repeat_region   6972. .7446
                     /note="L1MEc repeat: matches 278. .780 of consensus"
     repeat_region   7453. .8545
                     /note="HERVL repeat: matches 2899. .4004 of consensus"
     repeat_region   8794. .9044
                     /note="HERVL repeat: matches 2335. .2589 of consensus"
     repeat_region   9613. .9666
                     /note="Alu repeat: matches 1. .53 of consensus"
     repeat_region   9667. .9904
                     /note="MLT1D repeat: matches 168. .416 of consensus"
     repeat_region   9916. .10048
                     /note="AluJo repeat: matches 1. .136 of consensus"
     repeat_region   10884. .11146
                     /note="Tigger3(Golem) repeat: matches 2771. .3027 of
                     consensus"
     repeat_region   11160. .11284
                     /note="Tigger3(Golem) repeat: matches 1. .125 of consensus"
     repeat_region   11348. .11672
                     /note="AluJb repeat: matches 1. .310 of consensus"
     repeat_region   11707. .11794
                     /note="MIR repeat: matches 4. .92 of consensus"
     repeat_region   11798. .12173
                     /note="L1PA16 repeat: matches 5717. .6155 of consensus"
     repeat_region   12188. .12287
                     /note="MIR repeat: matches 138. .235 of consensus"
     repeat_region   12471. .16388
                     /note="L1P5 repeat: matches 2293. .6155 of consensus"
     repeat_region   16841. .16898
                     /note="L2 repeat: matches 2693. .2750 of consensus"
     repeat_region   18264. .18563
                     /note="AluSq repeat: matches 1. .300 of consensus"
     repeat_region   18637. .19137
                     /note="L1ME2 repeat: matches 5609. .6131 of consensus"
     repeat_region   19147. .19358
                     /note="53 copies 4 mer tata 75% conserved"
     repeat_region   19391. .19835
                     /note="L1ME2 repeat: matches 5135. .5581 of consensus"
     misc_feature    complement(19871. .20532)
                     /note="match: GSS: Em:AQ283704"
     repeat_region   19948. .20130
                     /note="THE1-INTERNAL repeat: matches 1398. .1580 of
                     consensus"
     repeat_region   20153. .20452
                     /note="THE1C repeat: matches 79. .371 of consensus"
     repeat_region   20453. .21357
                     /note="L1P repeat: matches 2739. .3657 of consensus"
     repeat_region   21330. .21610
                     /note="L1M4 repeat: matches 4033. .4299 of consensus"
     repeat_region   21611. .21907
                     /note="AluY repeat: matches 1. .298 of consensus"
     repeat_region   21908. .22023
                     /note="L1M4 repeat: matches 3915. .4033 of consensus"
     repeat_region   22053. .22549
```

```
                     /note="LTR44 repeat: matches 1. .519 of consensus"
     repeat_region   22814. .22841
                     /note="14 copies 2 mer tt 92% conserved"
     repeat_region   22922. .22969
                     /note="24 copies 2 mer tt 79% conserved"
     misc_feature    24230. .24396
                     /note="match: GSS: Em:AQ605495"
     repeat_region   24332. .24673
                     /note="MER65-internal repeat: matches 2142. .2492 of
                     consensus"
     repeat_region   24674. .24740
                     /note="MER51-internal repeat: matches 1581. .1971 of
                     consensus"
     repeat_region   25057. .25536
                     /note="L1PA5 repeat: matches 5655. .6143 of consensus"
     repeat_region   25537. .25600
                     /note="32 copies 2 mer aa 82% conserved"
     repeat_region   25625. .27912
                     /note="MER31-internal repeat: matches 3177. .4880 of
                     consensus"
     repeat_region   28380. .28818
                     /note="LTR44 repeat: matches 25. .519 of consensus"
     repeat_region   28830. .29139
                     /note="AluJb repeat: matches 1. .308 of consensus"
     repeat_region   29145. .30686
                     /note="L1M4 repeat: matches 2328. .3854 of consensus"
     repeat_region   30748. .31111
                     /note="THE1B repeat: matches 1. .364 of consensus"
     repeat_region   31112. .32639
                     /note="THE1B-INTERNAL repeat: matches 1. .1580 of
                     consensus"
     repeat_region   32640. .33010
                     /note="THE1B repeat: matches 1. .364 of consensus"
     repeat_region   33591. .33698
                     /note="MIR repeat: matches 33. .141 of consensus"
     repeat_region   34045. .34274
                     /note="MIR repeat: matches 2. .262 of consensus"
     repeat_region   35105. .35568
                     /note="MLT1C repeat: matches 1. .466 of consensus"
     repeat_region   36939. .37047
                     /note="Tigger3(Golem) repeat: matches 1. .108 of consensus"
     repeat_region   37039. .37745
                     /note="Tigger3(Golem) repeat: matches 2277. .3027 of
                     consensus"
     repeat_region   38735. .39129
                     /note="L2 repeat: matches 1162. .1576 of consensus"
     repeat_region   41401. .41633
                     /note="MER58A repeat: matches 1. .224 of consensus"
     repeat_region   42037. .42383
                     /note="LTR37A repeat: matches 76. .414 of consensus"
     misc_feature    complement(44109. .44432)
                     /note="match: GSS: Em:B35062"
     misc_feature    complement(44858. .45350)
                     /note="match: GSS: Em:AQ528681"
     misc_feature    complement(45342. .45500)
                     /note="match: GSS: Em:AQ149065"
     repeat_region   45661. .46730
                     /note="MER11C repeat: matches 1. .1071 of consensus"
     repeat_region   47237. .47615
                     /note="MSTA repeat: matches 1. .426 of consensus"
     misc_feature    47669. .48158
                     /note="match: GSS: Em:AQ428522"
     repeat_region   47693. .48244
                     /note="MER41A repeat: matches 1. .554 of consensus"
     repeat_region   48262. .48461
                     /note="MIR repeat: matches 53. .262 of consensus"
     misc_feature    48927. .49106
                     /note="match: GSS: Em:AQ075808"
     repeat_region   49149. .49332
                     /note="MIR repeat: matches 4. .192 of consensus"
     repeat_region   49401. .49551
                     /note="L1PA5 repeat: matches 5992. .6143 of consensus"
     repeat_region   50633. .50779
                     /note="L2 repeat: matches 2541. .2703 of consensus"
     repeat_region   51198. .51291
                     /note="MIR repeat: matches 14. .115 of consensus"

  Query Match              9.6%;  Score 192;  DB 9;  Length 204158;
  Best Local Similarity   73.0%;  Pred. No. 3.9e-34;
  Matches  305;  Conservative    0;  Mismatches  100;  Indels   13;  Gaps    4;

Qy        170 TTATTTACAGTCATGTGCCACATAACAATGTTTCAGTCAGGGATAG--AACACAAATGTA 227
              ||   |||||||||||||  |||||||  |||||||||  |||  |   | |||  |||
Db      11291 TTTAATACAGTCATGTGCTGCATAACACCATTTCAGTCAATGATGGGCTATACATATGAC 11232

Qy        228 TCTGGCCCCATAATATTATAAG----CTGAGAAATTTCTATTAACTAGTGATATCGCAGC 283
              | ||| | ||||| |||||||     |||| |||||  ||    |||||||  || | |||
Db      11231 TGTGGTCTCATAAAATTATAATGGAACTGAAAAATTCTTAAGGCCTAGTGACATTGTAGC 11172

Qy        284 CATCATAAG-TGTAATGCAGGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAA 342
              |||  |||  |    | ||   |||||||| ||| ||||||||||  ||||  |||||||
Db      11171 CATGGTAACCTCCTTTTCATAGCATTACCTGTTCGATGTTTAGATAGTTTAGGTACACAA 11112

Qy        343 ATATATTTCATTGTGTTATAATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGT 402
              |||  |  ||| |||||  || | |||||||  | |||||||  ||||||||||||||||
Db      11111 GTATTTACCATAGTGTTACAACTGCCTACAGAACTCAGTACTGTAACATGCTGTACAGGT 11052

Qy        403 TTGTAACCTAGGAGTAATAGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTAT----- 457
              | | | ||||  || |||||||  | ||||||    | |||||||  |||||||  ||
Db      11051 TGGCACCCTACCAGCAATAGGCCACACCATATGCCCTAGGTGTATAGTAGGGCATCTAGT 10992

Qy        458 -AACCATCTAGGTTTGTGTAAGTACATTCTATGATATTCCCACAATGATGAAATCACCTA 516
               ||||||||| |||||||||| |||||||||||| ||   |  |||||  ||||||||||
Db      10991 ACACCATCTAGATTTGTGTAAGGACATTCTATGATGTTAACGTAATGACAAAATCACCTA 10932

Qy        517 ACTACACATTTCTCAGAATGTTTCACTGTTGTGAAGTGACCCATGACTATATTTTCCT 574
              |  |||| |||||||||  | |  |  |||| |||||||  | |||||||||  | | |
Db      10931 ATGACACGTTTCTCAGAACATATTCCCATTGTTAAGTGACACCTGACTATACATCCAT 10874

RESULT 13
AC091607/c
LOCUS       AC091607               175447 bp    DNA     linear   PRI 31-JUL-2002
DEFINITION  Homo sapiens 3 BAC RP11-118D22 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC091607
VERSION     AC091607.11  GI:20986401
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175447)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
```

Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175447)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 175447)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 175447)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 175447)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (31-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 20, 2002 this sequence version replaced gi:20900906.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT.

FEATURES	Location/Qualifiers
source	1. .175447
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="3"
	/clone="RP11-118D22"
misc_feature	1. .2004
	/note="overlaps bases 28102. .30105 of clone AC125607"
	/function="clone overlap"
repeat_region	8. .44
	/rpt_family="CT-rich"
repeat_region	complement(310. .624)
	/rpt_family="AluY"
repeat_region	complement(1313. .1614)
	/rpt_family="AluSq"
repeat_region	3179. .3490
	/rpt_family="AluSx"
repeat_region	3491. .3541
	/rpt_family="(GGAA)n"
repeat_region	3756. .4049
	/rpt_family="L1ME"
repeat_region	4108. .4228
	/rpt_family="L1MEc"
repeat_region	4286. .4605
	/rpt_family="L2"
repeat_region	4629. .4652
	/rpt_family="AT_rich"
repeat_region	4664. .4885
	/rpt_family="L1MA4"
repeat_region	4906. .5644
	/rpt_family="Charlie2"
repeat_region	complement(5645. .6029)
	/rpt_family="MER7A"
repeat_region	6030. .6140
	/rpt_family="Charlie2"
repeat_region	complement(6260. .6744)
	/rpt_family="MLT1G"
repeat_region	complement(6746. .6889)
	/rpt_family="MIR"
repeat_region	8259. .8280
	/rpt_family="AT_rich"
repeat_region	complement(12402. .12711)
	/rpt_family="AluSx"
repeat_region	12753. .12815
	/rpt_family="(TA)n"
repeat_region	12971. .13031
	/rpt_family="MIR"
repeat_region	13053. .13346
	/rpt_family="AluSg"
repeat_region	14783. .14858
	/rpt_family="MIR"
repeat_region	14881. .15067
	/rpt_family="MIR"
repeat_region	15177. .15274
	/rpt_family="MIR"
repeat_region	16080. .16124
	/rpt_family="(TG)n"
repeat_region	17710. .17915
	/rpt_family="MIR"
repeat_region	18184. .18369
	/rpt_family="MIR"
repeat_region	complement(18598. .18814)

```
                     /rpt_family="MIR"
     repeat_region   19832. .19951
                     /rpt_family="FLAM_A"
     repeat_region   complement(20006. .20291)
                     /rpt_family="MLT1J2"
     repeat_region   complement(23184. .23491)
                     /rpt_family="AluSx"
     repeat_region   24263. .24367
                     /rpt_family="MER91C"
     repeat_region   complement(24408. .24629)
                     /rpt_family="MIR"
     repeat_region   complement(25224. .25366)
                     /rpt_family="L1PA13"
     repeat_region   complement(25748. .26050)
                     /rpt_family="AluSx"
     repeat_region   complement(26969. .27012)
                     /rpt_family="MIR"
     repeat_region   27780. .27805
                     /rpt_family="(T)n"
     repeat_region   complement(28026. .28071)

  Query Match             9.6%;  Score 191.6;  DB 9;  Length 175447;
  Best Local Similarity   72.0%;  Pred. No. 4.9e-34;
  Matches  324;  Conservative    0;  Mismatches  109;  Indels   17;  Gaps    5;

Qy        130 TTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTCATGTGCCA 189
              ||  |  || |||    |   | ||  |||| |||| ||    |    ||||||| ||||
Db       6077 TTCAACACTTTTAGGTATACAAACAACATTTTTCATCATAAATTAAGTAGTCATGAGCCA 6018

Qy        190 CATAACAATGTTTCAGTC--AGGGATAGAACACAAATGTATCTGGCCCCATAATATTATA 247
              |||||  |  |||||||||    |  |||   || | ||      ||| ||||||| ||||||
Db       6017 CATAATGACATTTCAGTCGATGACATACTGCATACATCATGGTGGTCCCATAAGATTATA 5958

Qy        248 A----GCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCAGG 303
              |    ||||| |||||| ||| |||||||||| |||  ||  |||||||    |   | |
Db       5957 ATGGTGCTGAAAAATTTTTATAAACTAGTGACATCATAGTTATCATAATGTTGTAGTACA 5898

Qy        304 A-CATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTATA 362
              | ||||||||||||| | |         |||||||||||||||  |  ||||||||||||
Db       5897 ACCATTACCTTTTCTGTAT---------TTAGATACACAAATACTTACCATTGTGTTATA 5847

Qy        363 ATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAG 422
              ||| |||||| |||| ||||||| || ||||| || ||||||||| | ||| || ||||
Db       5846 ATTGCCTACAATATTTAGTACAGCAATATGCTATAAAGGTTTGTAGCTTAGTAGCAATAA 5787

Qy        423 GCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACA 482
              ||| ||||||| ||| ||||| || | |||||||| ||||||||||||||||||||| ||
Db       5786 GCTCTACCATATAGCCTAGGTATGGAATAGGCTAT-ACCATCTAGGTTTGTGTAAGTGCA 5728

Qy        483 TTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCAC 542
              |||||||||||||||||| ||||||| ||||| |  |||||||||| ||  | |
Db       5727 TTCTATGATATTCCCACAAGTATGAAATTGCCTAAGGAGGCATTTCTCAGAAAGTAACCC 5668

Qy        543 TGTTGTGAAGTGACCCATGACTATATTTTC 572
              |||||| |||  |  ||  ||| || || |
Db       5667 TGTTGTTAAGGAATGCACCACTGTAGTTGC 5638


RESULT 14
AC024160
LOCUS       AC024160               176417 bp    DNA     linear   HTG 29-MAY-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-294L13 map 3p, WORKING DRAFT
            SEQUENCE, 19 unordered pieces.
ACCESSION   AC024160
VERSION     AC024160.3  GI:8101179
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176417)
  AUTHORS   Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
            Dong,W., Fan,H., Feng,X., Guan,Q., Gu,X., Guo,D., He,L., Hu,S.,
            Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,G., Li,J., Li,L.,
            Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y., Li,W., Li,W., Li,Y.,
            Luo,J., Niu,Y., Qi,Q., Qi,X., Song,S., Sun,M., Sun,W., Sun,Y.,
            Tao,R., Wang,H., Wang,J., Wang,J., Wang,L., Wang,L., Wang,R.,
            Wang,X., Wang,X., Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y.,
            Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H.,
            Zhang,L., Zhang,M., Zhang,X., Zhang,X., Zhang,Y., Zhang,Y.,
            Zhang,Z., Zhu,B., Yu,J. and Yang,H.
  TITLE     Chromosome 3p genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176417)
  AUTHORS   Chen,C., Hu,S., Dong,W., Zhang,X., Wang,J., Zhang,Y., Zhang,H.,
            Liu,B., Bao,W., Sun,Y., Wu,Q., Wang,H., Yang,X., Cheng,C., Wang,Y.,
            Niu,Y., Qi,X., Li,T., Zhang,H., Liu,N., Wu,D., Yu,B., Fan,H.,
            Liu,Y., Li,G., Li,C., Bao,Q., Bao,J., Wang,X., Song,L., Zhang,L.,
            Guo,D., Huang,F., Zhang,G., Li,J., Bian,X., Zhang,M., Li,L.,
            Feng,X., Yu,J. and Yang,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2000) Human Genomic Center, Institute of
            Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
            100101, P.R.China
COMMENT     On May 29, 2000 this sequence version replaced gi:7644436.
            --------------Genome Center
            Center:Beijing Center
            Center code:Beijing
            Website:http://hgc.igtp.ac.cn
            http://www.genomics.org.cn
            Contact:hgc@igtp.ac.cn
            -------------- Project Information
            Center project name:1% project
            Center clone name: RP11-294L13
            -------------- Summary Statistics
            Sequencing vector: pUC18; 100% of reads
            Chemistry: Dye-terminator: ET 55% of reads
            Chemistry: Dye-terminator Big Dye; 45% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 165854 bases at least Q40
            Consensus quality: 170551 bases at least Q30
            Consensus quality: 173535 bases at least Q20
            Insert size: 167336; sum-of-contigs
            Quality coverage:   4.26x in Q20 bases;sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1505: contig of 1505 bp in length
            *     1506     1605: gap of unknown length
            *     1606     4034: contig of 2429 bp in length
            *     4035     4134: gap of unknown length
            *     4135     5856: contig of 1722 bp in length
            *     5857     5956: gap of unknown length
            *     5957     8671: contig of 2715 bp in length
            *     8672     8771: gap of unknown length
            *     8772    14318: contig of 5547 bp in length
            *    14319    14418: gap of unknown length
            *    14419    18764: contig of 4346 bp in length
            *    18765    18864: gap of unknown length
            *    18865    23344: contig of 4480 bp in length
            *    23345    23444: gap of unknown length
            *    23445    28654: contig of 5210 bp in length
            *    28655    28754: gap of unknown length
            *    28755    34797: contig of 6043 bp in length
            *    34798    34897: gap of unknown length
            *    34898    40772: contig of 5875 bp in length
            *    40773    40872: gap of unknown length
            *    40873    47168: contig of 6296 bp in length
            *    47169    47268: gap of unknown length
            *    47269    53552: contig of 6284 bp in length
```

```
            *    53553    53652: gap of unknown length
            *    53653    65854: contig of 12202 bp in length
            *    65855    65954: gap of unknown length
            *    65955    80747: contig of 14793 bp in length
            *    80748    80847: gap of unknown length
            *    80848    92030: contig of 11183 bp in length
            *    92031    92130: gap of unknown length
            *    92131   109264: contig of 17134 bp in length
            *   109265   109364: gap of unknown length
            *   109365   123322: contig of 13958 bp in length
            *   123323   123422: gap of unknown length
            *   123423   146425: contig of 23003 bp in length
            *   146426   146525: gap of unknown length
            *   146526   176417: contig of 29892 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176417
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3p"
                     /clone="RP11-294L13"
     misc_feature    1. .1505
                     /note="assembly_name:Contig3"
     misc_feature    1606. .4034
                     /note="assembly_name:Contig5"
     misc_feature    4135. .5856
                     /note="assembly_name:Contig6"
     misc_feature    5957. .8671
                     /note="assembly_name:Contig7"
     misc_feature    8772. .14318
                     /note="assembly_name:Contig8"
     misc_feature    14419. .18764
                     /note="assembly_name:Contig9"
     misc_feature    18865. .23344
                     /note="assembly_name:Contig10"
     misc_feature    23445. .28654
                     /note="assembly_name:Contig11"
     misc_feature    28755. .34797
                     /note="assembly_name:Contig12"
     misc_feature    34898. .40772
                     /note="assembly_name:Contig13"
     misc_feature    40873. .47168
                     /note="assembly_name:Contig14"
     misc_feature    47269. .53552
                     /note="assembly_name:Contig15"
     misc_feature    53653. .65854
                     /note="assembly_name:Contig16"
     misc_feature    65955. .80747
                     /note="assembly_name:Contig17"
     misc_feature    80848. .92030
                     /note="assembly_name:Contig18"
     misc_feature    92131. .109264
                     /note="assembly_name:Contig19"
     misc_feature    109365. .123322
                     /note="assembly_name:Contig20"
     misc_feature    123423. .146425
                     /note="assembly_name:Contig21
                     clone_end:SP6
                     vector_side:right"
     misc_feature    146526. .176417
                     /note="assembly_name:Contig22"
BASE COUNT    51307 a  35424 c  35453 g  52378 t   1855 others
ORIGIN

  Query Match             9.6%;  Score 191.6;  DB 2;  Length 176417;
  Best Local Similarity   72.0%;  Pred. No. 4.9e-34;
  Matches  324;  Conservative    0;  Mismatches  109;  Indels   17;  Gaps    5;

Qy        130 TTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTCATGTGCCA 189
              ||  |  || |||    |   | ||  |||| |||| ||    |    ||||||| ||||
Db     150340 TTCAACACTTTTAGGTATACAAACAACATTTTTCATCATAAATTAAGTAGTCATGAGCCA 150399

Qy        190 CATAACAATGTTTCAGTC--AGGGATAGAACACAAATGTATCTGGCCCCATAATATTATA 247
              |||||  |  ||||||||    |  |||   || | ||     ||| |||||||  ||||||
Db     150400 CATAATGACATTTCAGTCGATGACATACTGCATACATCATGGTGGTCCCATAAGATTATA 150459

Qy        248 A----GCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCAGG 303
              |    ||||| |||||| ||| |||||||| |||  ||  |||||||   |   | |
Db     150460 ATGGTGCTGAAAAATTTTTATAAACTAGTGACATCATAGTTATCATAATGTTGTAGTACA 150519

Qy        304 A-CATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTATA 362
              | ||||||||||||| | |         ||||||||||||||  |  ||||||||||||
Db     150520 ACCATTACCTTTTCTGTAT---------TTAGATACACAAATACTTACCATTGTGTTATA 150570

Qy        363 ATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAG 422
              ||| |||||| |||| ||||||| || ||||| || |||||||| | ||| || ||||
Db     150571 ATTGCCTACAATATTTAGTACAGCAATATGCTATAAAGGTTTGTAGCTTAGTAGCAATAA 150630

Qy        423 GCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACA 482
              ||| ||||||| ||| ||||| || | |||||||| ||||||||||||||||||||| ||
Db     150631 GCTCTACCATATAGCCTAGGTATGGAATAGGCTAT-ACCATCTAGGTTTGTGTAAGTGCA 150689

Qy        483 TTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCAC 542
              |||||||||||||||||  |||||||  |||||  |  |||||||||| ||  | |
Db     150690 TTCTATGATATTCCCACAAGTATGAAATTGCCTAAGGAGGCATTTCTCAGAAAGTAACCC 150749

Qy        543 TGTTGTGAAGTGACCCATGACTATATTTTC 572
              |||||| |||  |  ||  ||| || || |
Db     150750 TGTTGTTAAGGAATGCACCACTGTAGTTGC 150779


RESULT 15
AL592156/c
LOCUS       AL592156               134995 bp    DNA     linear   PRI 20-JUL-2001
DEFINITION  Human DNA sequence from clone RP11-492O8 on chromosome X, complete
            sequence.
ACCESSION   AL592156
VERSION     AL592156.4  GI:15020776
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 134995)
  AUTHORS   Wilson,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 25, 2001 this sequence version replaced gi:14625708.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP11-492O8 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
```

```
          VECTOR: pBACe3.6
          IMPORTANT: This sequence is not the entire insert of clone
          RP11-492O8 It may be shorter because we sequence overlapping
          sections only once, except for a 100 base overlap.
          The true right end of clone RP11-492O8 is at 134995 in this
          sequence. The true right end of clone RP13-46M24 is at 2000 in this
          sequence.
FEATURES             Location/Qualifiers
     source          1. .134995
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-492O8"
                     /clone_lib="RPCI-11.2"
     repeat_region   1. .925
                     /note="L1MEc repeat: matches 345. .1276 of consensus"
     repeat_region   1197. .1351
                     /note="MIR repeat: matches 87. .248 of consensus"
     repeat_region   3508. .3643
                     /note="2 copies 68 mer 86% conserved"
     repeat_region   3787. .3964
                     /note="LTR31 repeat: matches 418. .603 of consensus"
     repeat_region   3816. .3966
                     /note="LTR26 repeat: matches 433. .591 of consensus"
     repeat_region   4366. .4440
                     /note="MER39 repeat: matches 13. .84 of consensus"
     repeat_region   4441. .4735
                     /note="AluJo repeat: matches 1. .294 of consensus"
     repeat_region   4736. .5023
                     /note="MER39 repeat: matches 84. .387 of consensus"
     repeat_region   5900. .6050
                     /note="LTR33 repeat: matches 36. .217 of consensus"
     repeat_region   9982. .10500
                     /note="MER70B repeat: matches 127. .653 of consensus"
     repeat_region   10958. .11035
                     /note="L1PA4 repeat: matches 5701. .5782 of consensus"
     repeat_region   11036. .11399
                     /note="L1PA4 repeat: matches 5783. .6146 of consensus"
     repeat_region   11408. .11447
                     /note="20 copies 2 mer aa 85% conserved"
     repeat_region   11663. .12394
                     /note="L1PA13 repeat: matches -651. .106 of consensus"
     repeat_region   12445. .15153
                     /note="L1PA13 repeat: matches 653. .3418 of consensus"
     repeat_region   15157. .15448
                     /note="AluSx repeat: matches 2. .303 of consensus"
     repeat_region   17246. .17405
                     /note="MER69A repeat: matches 5. .168 of consensus"
     repeat_region   17923. .17986
                     /note="MIR repeat: matches 27. .89 of consensus"
     repeat_region   18336. .18619
                     /note="AluJb repeat: matches 1. .287 of consensus"
     repeat_region   19591. .20009
                     /note="L1ME repeat: matches 5530. .5966 of consensus"
     repeat_region   20098. .21129
                     /note="L1ME repeat: matches 4264. .5398 of consensus"
     repeat_region   21295. .21707
                     /note="MLT1A1 repeat: matches 3. .392 of consensus"
     repeat_region   21813. .21887
                     /note="L1M4 repeat: matches 3830. .3909 of consensus"
     repeat_region   22685. .22768
                     /note="42 copies 2 mer tt 64% conserved"
     repeat_region   23023. .23070
                     /note="L1MA4 repeat: matches 5944. .5992 of consensus"
     repeat_region   23071. .23375
                     /note="AluSq repeat: matches 1. .304 of consensus"
     repeat_region   23376. .23511
                     /note="L1MA4 repeat: matches 5814. .5944 of consensus"
     repeat_region   23530. .25674
                     /note="L1MA2 repeat: matches 4120. .6308 of consensus"
     repeat_region   25696. .25795
                     /note="L1 repeat: matches 4067. .4156 of consensus"
     repeat_region   25796. .25965
                     /note="MSTA repeat: matches 216. .426 of consensus"
     repeat_region   25966. .26152
                     /note="L1MA1 repeat: matches 5971. .6162 of consensus"
     repeat_region   26154. .26636
                     /note="L1 repeat: matches 3540. .4022 of consensus"
     repeat_region   26637. .26774
                     /note="L1PA7 repeat: matches 6006. .6143 of consensus"
     repeat_region   26775. .28021
                     /note="L1 repeat: matches 2293. .3540 of consensus"
     repeat_region   28038. .28192
                     /note="L1MA1 repeat: matches 6136. .6293 of consensus"
     repeat_region   28193. .28577
                     /note="MSTA repeat: matches 1. .426 of consensus"
     repeat_region   28578. .28893
                     /note="L1MA1 repeat: matches 5834. .6136 of consensus"
     repeat_region   28894. .32496
                     /note="L1M1 repeat: matches -1389. .2667 of consensus"
     repeat_region   32509. .34531
                     /note="L1MA4 repeat: matches 3784. .5828 of consensus"
     repeat_region   34804. .35020
                     /note="LTR16A repeat: matches 150. .366 of consensus"
     repeat_region   35092. .35286
                     /note="HERV16 repeat: matches 1. .208 of consensus"
     repeat_region   35289. .35661
                     /note="MER1B repeat: matches 1. .332 of consensus"
     repeat_region   35855. .36063
                     /note="HERV16 repeat: matches 424. .639 of consensus"
     repeat_region   36418. .36707
                     /note="HERVL repeat: matches 3350. .3654 of consensus"
     repeat_region   36736. .37517
                     /note="MER21B repeat: matches 1. .790 of consensus"
     repeat_region   37694. .38477
                     /note="HERVL repeat: matches 3880. .4688 of consensus"
     repeat_region   39147. .39497
                     /note="LTR33 repeat: matches 171. .521 of consensus"
     repeat_region   39507. .40018
                     /note="L1MB7 repeat: matches 5644. .6173 of consensus"
     repeat_region   40032. .40208
                     /note="L1MB7 repeat: matches 5353. .5531 of consensus"
     repeat_region   40218. .40661
                     /note="L1MA1 repeat: matches 5861. .6304 of consensus"
     repeat_region   40685. .41473
                     /note="L1M1 repeat: matches -1389. .-542 of consensus"
     repeat_region   41484. .42354
                     /note="L1M4 repeat: matches 3522. .4414 of consensus"
     repeat_region   42352. .43064
                     /note="L1MA8 repeat: matches 5476. .6177 of consensus"
     repeat_region   43165. .43192
                     /note="L1MA8 repeat: matches 5444. .5471 of consensus"
     repeat_region   43193. .43281
                     /note="L1MA4 repeat: matches 6210. .6297 of consensus"
     repeat_region   43282. .43400
                     /note="L1MA8 repeat: matches 5354. .5445 of consensus"
     repeat_region   44383. .44495
                     /note="LTR9 repeat: matches 302. .415 of consensus"
     repeat_region   44469. .44740
                     /note="LTR9 repeat: matches 346. .625 of consensus"
     repeat_region   44817. .44917
                     /note="MSTA repeat: matches 1. .99 of consensus"
     repeat_region   44918. .45282
                     /note="THE1B repeat: matches 1. .364 of consensus"
     repeat_region   45283. .45601
                     /note="MSTA repeat: matches 99. .426 of consensus"
     repeat_region   47118. .47737
                     /note="HAL1 repeat: matches 324. .995 of consensus"
     repeat_region   47995. .48197
                     /note="L2 repeat: matches 2549. .2750 of consensus"
     repeat_region   48206. .48495
                     /note="L1MA3 repeat: matches 6027. .6304 of consensus"
     repeat_region   50492. .50829
                     /note="THE1C repeat: matches 1. .371 of consensus"
     repeat_region   51649. .52043
```

```
                     /note="MER39 repeat: matches 31. .443 of consensus"
     repeat_region   52235. .52281
                     /note="MER5A repeat: matches 143. .189 of consensus"
     repeat_region   52379. .52607
                     /note="L1PA7 repeat: matches 5898. .6142 of consensus"
     repeat_region   52612. .52799
                     /note="L1PA16 repeat: matches 5963. .6156 of consensus"
     repeat_region   53636. .53810
                     /note="L1MA4 repeat: matches 6112. .6300 of consensus"
     repeat_region   53765. .54022
                     /note="L1 repeat: matches 3324. .3580 of consensus"
     repeat_region   54024. .54176

  Query Match             9.5%;  Score 191;  DB 9;  Length 134995;
  Best Local Similarity   77.0%;  Pred. No. 7e-34;
  Matches  312;  Conservative    0;  Mismatches   80;  Indels   13;  Gaps    6;

Qy        169 TTTATTTACAGTCATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTAT 228
              | | | ||| ||||| |||||||||||  |||||  ||||  | | || |||| || |
Db     109919 TATGTGTACCGTCATATGCCACATAACGGTGTTTTGGTCAATGGTGGACCACATATATGG 109860

Qy        229 C--TGGCCCCATAATATTATA----AGCTGAGAAATTTCTATTAACTAGTGATATCGCA- 281
              |  |||   | ||| || |||    |||||| ||||||||||  |||||||| | |
Db     109859 CAGTGGTTGCGTAAAATCATAATGGAGCTGAAAAATTTCTATTGCCTAGTGATGTTGTGG 109800

Qy        282 --GCCATCATAAGTGTAATGCAGGACATTACCTTTTCTATGTTTAGATATG-TTAGATAC 338
                |||  ||||||   |   |    |||||||||||||||| | ||||||||  ||||||||
Db     109799 TGGCCCTCATAACATTGTAGTGCAACATTACCTTTTCTATATATAGATATGTTTAGATAC 109740

Qy        339 ACAAATATATTTCATTGTGTTATAATTTCCTACAGTATTCAGTACAGTAACATGCTGTAC 398
              |||||||  | |||||||  ||| ||||  ||||||  |||||  |||  ||||||||||||||||
Db     109739 ACAAATACTTATCATTGTATTACAATTACCTACAATATTCTGTATAGTAACATGCTGTAC 109680

Qy        399 AGGTTTGTAACCTAGGAGTAATAGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATA 458
              |||| |||| |||||||| |||||||| |||||| ||| | |||||||| |||||||
Db     109679 AGGTATGTAGCCTAGGAGAAATAGGCTACACCATATAGCCTGGGTGTGTAGCAGGCTAT- 109621

Qy        459 ACCATCTAGGTTTGTGTAAGTACATTCTATGATAT--TCCCACAATGATGAAATCACCTA 516
              |||||||||| ||||||||||   ||||||||||  || || ||  |  ||||||||||
Db     109620 ACCATCTAGGTCTGTGTAAGTATGCTCTATGATATGGTCGCATAACAACAAAATCACCTA 109561

Qy        517 ACTACACATTTCTCAGAATGTTTCACTGTTGTGAAGTGACCCATG 561
              || ||||||||||||||||  |||| |  |  |||||| ||||
Db     109560 ACAACACATTTCTCAGAATGTACCACTATCATTCAGTGACACATG 109516


Search completed: November  7, 2003, 11:11:12
Job time : 7845.84 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 01:54:16 ; Search time 1962.4 Seconds
(without alignments)
10444.222 Million cell updates/sec

Title: US-09-939-209A-3_COPY_10000_10500
Perfect score: 501
Sequence: 1 gaccatgtataatatgatgc.....cgctcaaagccgaagccaca 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	501	100.0	20300	6	AX451337	AX451337 Sequence
	2	501	100.0	165329	9	AL583850	AL583850 Human DNA
c	3	501	100.0	191699	2	AC031977	AC031977 Homo sapi
c	4	167.8	33.5	207387	2	AC115766	AC115766 Mus muscu
c	5	138.2	27.6	275631	2	AC125563	AC125563 Rattus no
	6	69.4	13.9	2599	9	AK093959	AK093959 Homo sapi
	7	69	13.8	3769	9	AK096204	AK096204 Homo sapi
c	8	59	11.8	234560	2	AC111539	AC111539 Rattus no
	9	48	9.6	174328	9	AC100845	AC100845 Homo sapi
	10	47.6	9.5	7218	6	I66494	I66494 Sequence 14
	11	45.6	9.1	307979	2	AC095234	AC095234 Rattus no
	12	45.4	9.1	10020	6	AX347195	AX347195 Sequence
c	13	45.4	9.1	134971	2	AC116367	AC116367 Oryza sat
	14	45.2	9.0	198464	2	AC022764	AC022764 Homo sapi
c	15	45.2	9.0	203581	10	AL451076	AL451076 Mouse DNA
c	16	45.2	9.0	252721	2	AC095946	AC095946 Rattus no
	17	44.8	8.9	1122	3	AF020286	AF020286 Dictyoste
	18	44	8.8	2753	9	BC051869	BC051869 Homo sapi
	19	43.6	8.7	144177	9	AL513487	AL513487 Human DNA
	20	43.4	8.7	119171	2	BX247904	BX247904 Danio rer
	21	43.4	8.7	124895	8	AC135288	AC135288 Solanum d
	22	43.2	8.6	179212	9	AC084877	AC084877 Homo sapi
c	23	43	8.6	59469	2	AC110024	AC110024 Homo sapi
c	24	43	8.6	173817	9	AC103834	AC103834 Homo sapi
	25	42.8	8.5	125020	9	AF429315	AF429315 Homo sapi
c	26	42.4	8.5	6210	9	AF053319	AF053319 Homo sapi
	27	42.4	8.5	110000	9	AE014305_1	Continuation (2 of
	28	42.4	8.5	165916	9	AC112203	AC112203 Homo sapi
	29	42.4	8.5	167803	2	AC010816	AC010816 Homo sapi
c	30	42.2	8.4	59584	2	AC090243	AC090243 Homo sapi
	31	42.2	8.4	166315	9	AC078940	AC078940 Homo sapi
c	32	42	8.4	887	9	HSM801336	AL133065 Homo sapi
	33	42	8.4	83912	8	AP002043	AP002043 Arabidops
c	34	42	8.4	131275	10	AL669860	AL669860 Mouse DNA
c	35	42	8.4	150681	9	AP001251	AP001251 Homo sapi
c	36	42	8.4	192187	3	AC116920	AC116920 Dictyoste
c	37	42	8.4	201299	5	AL773601	AL773601 Zebrafish
c	38	42	8.4	250029	3	AE014839	AE014839 Plasmodiu
c	39	42	8.4	340000	9	AP001679	AP001679 Homo sapi
	40	41.6	8.3	146570	3	AC117072	AC117072 Dictyoste
	41	41.6	8.3	150351	2	AC102755	AC102755 Mus muscu
	42	41.6	8.3	250029	3	AE014820	AE014820 Plasmodiu
c	43	41.4	8.3	38692	3	AC116919	AC116919 Dictyoste
	44	41.4	8.3	100549	9	AL138776	AL138776 Human DNA
c	45	41.4	8.3	110000	2	PFMAL7P1_08	Continuation (9 of

ALIGNMENTS

RESULT 1
AX451337
LOCUS AX451337 20300 bp DNA linear PAT 03-JUL-2002
DEFINITION Sequence 3 from Patent WO0216653.
ACCESSION AX451337
VERSION AX451337.1 GI:21698388
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Levitt,P.R., Mirnics,K., Kodavali,V.C. and Nimgaonkar,V.L.
TITLE Methods and systems for facilitating the diagnosis and treatment of schizophrenia
JOURNAL Patent: WO 0216653-A 3 28-FEB-2002;

```
            University of Pittsburgh (US)
FEATURES             Location/Qualifiers
     source          1. .20300
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="A genomic sequence containing RGS4 nucleic acid
                     sequence and sequ ences upstream and downstream to the
                     RGS4 nucleic acid sequence"
BASE COUNT     6157 a    4102 c    3775 g    6266 t
ORIGIN

  Query Match             100.0%;  Score 501;  DB 6;  Length 20300;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-113;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GACCATGTATAATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10000 GACCATGTATAATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCT 10059

Qy         61 AAGTAAGCTCCAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10060 AAGTAAGCTCCAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATT 10119

Qy        121 TTTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10120 TTTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCT 10179

Qy        181 TTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10180 TTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATAT 10239

Qy        241 CCCTACTTTTCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10240 CCCTACTTTTCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCT 10299

Qy        301 TTTCTTCCTCATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10300 TTTCTTCCTCATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTG 10359

Qy        361 CAGAGCGGTCGTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10360 CAGAGCGGTCGTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAG 10419

Qy        421 GCTTAGCAGGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10420 GCTTAGCAGGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGT 10479

Qy        481 ACGCTCAAAGCCGAAGCCACA 501
              |||||||||||||||||||||
Db      10480 ACGCTCAAAGCCGAAGCCACA 10500


RESULT 2
AL583850
LOCUS       AL583850               165329 bp    DNA     linear   PRI 15-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-430G6 on chromosome 1, complete
            sequence.
ACCESSION   AL583850
VERSION     AL583850.5  GI:16973044
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165329)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 16, 2001 this sequence version replaced gi:15020514.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-430G6 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-430G6 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP11-430G6 is at 165329 in this
            sequence. The true right end of clone RP11-331H2 is at 2000 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .165329
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-430G6"
                     /clone_lib="RPCI-11.2"
BASE COUNT    51144 a  30897 c  31439 g  51849 t
ORIGIN

  Query Match             100.0%;  Score 501;  DB 9;  Length 165329;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-113;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GACCATGTATAATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81681 GACCATGTATAATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCT 81740

Qy         61 AAGTAAGCTCCAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81741 AAGTAAGCTCCAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATT 81800

Qy        121 TTTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81801 TTTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCT 81860

Qy        181 TTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81861 TTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATAT 81920

Qy        241 CCCTACTTTTCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81921 CCCTACTTTTCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCT 81980

Qy        301 TTTCTTCCTCATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81981 TTTCTTCCTCATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTG 82040

Qy        361 CAGAGCGGTCGTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82041 CAGAGCGGTCGTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAG 82100
```

```
Qy        421 GCTTAGCAGGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82101 GCTTAGCAGGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGT 82160

Qy        481 ACGCTCAAAGCCGAAGCCACA 501
              |||||||||||||||||||||
Db      82161 ACGCTCAAAGCCGAAGCCACA 82181


RESULT 3
AC031977/c
LOCUS       AC031977               191699 bp    DNA     linear   HTG 12-APR-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-288O18, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC031977
VERSION     AC031977.7  GI:13194952
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
            Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
            Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
            Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Mar 4, 2001 this sequence version replaced gi:9665085.
            -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 880
                 Center clone name: RP11-288O18
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513; 100% of reads
                 Sequencing Vector: plasmid; plasmid_accession; 0% of reads
                 Chemistry: Dye-primer; 1% of reads
                 Chemistry: Dye-terminator Big Dye; 99% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 190680 bases at least Q40
                 Consensus quality: 191287 bases at least Q30
                 Consensus quality: 191336 bases at least Q20
                 Insert size: 195548; agarose-fp
                 Insert size: 191499; sum-of-contigs
                 Quality coverage:  7.9x in Q20 bases; agarose-fp
                 Quality coverage:  8.1x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    12646: contig of 12646 bp in length
            *    12647    12746: gap of unknown length
            *    12747    94961: contig of 82215 bp in length
            *    94962    95061: gap of unknown length
            *    95062   191699: contig of 96638 bp in length.
FEATURES             Location/Qualifiers
     source          1. .191699
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-288O18"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .12646
                     /note="assembly_name:Contig5"
     misc_feature    12747. .94961
                     /note="assembly_name:Contig6
                     clone_end:T7"
     misc_feature    95062. .191699
                     /note="assembly_name:Contig7
                     clone_end:SP6"
BASE COUNT    57654 a  36385 c  36166 g  61293 t    201 others
ORIGIN

  Query Match           100.0%;  Score 501;  DB 2;  Length 191699;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-113;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GACCATGTATAATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2937 GACCATGTATAATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCT 2878

Qy         61 AAGTAAGCTCCAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2877 AAGTAAGCTCCAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATT 2818

Qy        121 TTTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2817 TTTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCT 2758

Qy        181 TTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2757 TTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATAT 2698

Qy        241 CCCTACTTTTCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2697 CCCTACTTTTCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCT 2638

Qy        301 TTTCTTCCTCATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2637 TTTCTTCCTCATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTG 2578

Qy        361 CAGAGCGGTCGTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2577 CAGAGCGGTCGTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAG 2518

Qy        421 GCTTAGCAGGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2517 GCTTAGCAGGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGT 2458

Qy        481 ACGCTCAAAGCCGAAGCCACA 501
              |||||||||||||||||||||
Db       2457 ACGCTCAAAGCCGAAGCCACA 2437


RESULT 4
AC115766/c
LOCUS       AC115766               207387 bp    DNA     linear   HTG 23-MAR-2003
DEFINITION  Mus musculus clone RP23-82I24, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   AC115766
VERSION     AC115766.4  GI:29164623
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 207387)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-82I24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 207387)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 207387)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 23, 2003 this sequence version replaced gi:28191504.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L23317
                Center clone name: 82_I_24
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 205497 bases at least Q40
                Consensus quality: 206243 bases at least Q30
                Consensus quality: 206432 bases at least Q20
                Insert size: 205000; agarose-fp
                Insert size: 206587; sum-of-contigs
                Quality coverage: 8.8 in Q20 bases; agarose-fp
                Quality coverage: 8.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    61925: contig of 61925 bp in length
            *    61926    62025: gap of 100 bp
            *    62026    64603: contig of 2578 bp in length
            *    64604    64703: gap of 100 bp
            *    64704    68576: contig of 3873 bp in length
            *    68577    68676: gap of 100 bp
            *    68677    71936: contig of 3260 bp in length
            *    71937    72036: gap of 100 bp
            *    72037    76569: contig of 4533 bp in length
            *    76570    76669: gap of 100 bp
            *    76670    82563: contig of 5894 bp in length
            *    82564    82663: gap of 100 bp
            *    82664   119650: contig of 36987 bp in length
            *   119651   119750: gap of 100 bp
            *   119751   165174: contig of 45424 bp in length
            *   165175   165274: gap of 100 bp
            *   165275   207387: contig of 42113 bp in length.
FEATURES             Location/Qualifiers
     source          1. .207387
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-82I24"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .61925
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    62026. .64603
                     /note="assembly_fragment"
     misc_feature    64704. .68576
                     /note="assembly_fragment"
     misc_feature    68677. .71936
                     /note="assembly_fragment"
     misc_feature    72037. .76569
                     /note="assembly_fragment"
     misc_feature    76670. .82563
                     /note="assembly_fragment"
     misc_feature    82664. .119650
                     /note="assembly_fragment"
     misc_feature    119751. .165174
                     /note="assembly_fragment"
     misc_feature    165275. .207387
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    60516 a  42224 c  43121 g  60726 t    800 others
ORIGIN

  Query Match           33.5%;  Score 167.8;  DB 2;  Length 207387;
  Best Local Similarity   63.8%;  Pred. No. 5.8e-31;
  Matches  307;  Conservative    0;  Mismatches  162;  Indels   12;  Gaps    3;

Qy         12 ATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCTAAGTAAGCTCC 71
              |  ||  ||||||  ||||      | | | |||| ||  |||      |||| | |
Db     115085 AGAATATATTTCTAACCCAAGGAGAAACATGTATTGAGACCCAGACGTAGAGTAGGATGA 115026

Qy         72 AGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTACA 131
              ||| | | |||||||    |||| |||  |||    | ||||||      ||  | ||   |
Db     115025 AGACTCCTTGCTGGTTTCTTTCTTTCTGGGAGAACAGTTCCTTAGACATTGATCTTACAA 114966
```

```
Qy        132 GGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAGAAAT 191
              | |          ||  | ||| |||      |  || |  || | ||||||| ||| ||||
Db     114965 GCC--------AATGAATATTTTTCAGCACGGTGGATTCTATATTTCCTTTCCTACAAAT 114914

Qy        192 TCTAAACCTCTGACATTGGTGGAGACATTGAG---TACATTTTTTCCCATATCCCTACTT 248
              | |||| |    | | ||| || || || ||     || |  ||| | |||| |||||||
Db     114913 GCAAAACTTGACGCGTAGGTAGAAACGTTAAGCAAAACCTCATTTTCTATATGCCTACTT 114854

Qy        249 TTCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCTTTTCTTCC 308
              | |  | || || |||| ||| | ||  ||  |||  ||||||||||||  |||||||
Db     114853 TGCGAATGGCTTCTCTCCGCTGGGTCCATTTGCATCCCTGATGCGTCAGTTGTTTCTTCC 114794

Qy        309 TCATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGG 368
              ||| |  ||    ||||||||||||||| | |    |||  |||| |   ||||||| | |
Db     114793 TCACCCTTTCTGAGGGCTGGAGAGGCAGCGCGCCGCAGCAGAGCGGAGCCTGCAGACCTG 114734

Qy        369 TCGTCTGATTGGCTGGACGGT-CGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGC 427
              || ||||||||||||   ||| ||| |||| ||||||||||||||||  | |||||
Db     114733 TCATCTGATTGGCTGCCGGGTGCGTGGCTGAGCTATAAAAGAGACCCCACCGGGCTTGAA 114674

Qy        428 AGGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCA 487
              ||  ||| ||||||||||||||||||   ||||| | |   |  ||  |  || | | |
Db     114673 AGCCAGAGGCTCAGAGGATTCTGACAGCTTCTTTGCAGAGCAGAAGCAACTGTGCTCAAA 114614

Qy        488 A 488
              |
Db     114613 A 114613
```

RESULT 5
AC125563/c
LOCUS AC125563 275631 bp DNA linear HTG 09-NOV-2002
DEFINITION Rattus norvegicus clone CH230-9B12, WORKING DRAFT SEQUENCE, 4 unordered pieces.
ACCESSION AC125563
VERSION AC125563.4 GI:24817906
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 275631)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 275631)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 275631)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 9, 2002 this sequence version replaced gi:23096552.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GDBK
Center clone name: CH230-9B12
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 228152 bases at least Q40
Consensus quality: 230310 bases at least Q30
Consensus quality: 231872 bases at least Q20
Estimated insert size: 235029; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
            *       1   192878: contig of 192878 bp in length
            *  192879   192978: gap of unknown length
            *  192979   270295: contig of 77317 bp in length
            *  270296   270395: gap of unknown length
            *  270396   272405: contig of 2010 bp in length
            *  272406   272505: gap of unknown length
            *  272506   275631: contig of 3126 bp in length.
FEATURES             Location/Qualifiers
     source          1. .275631
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-9B12"
     misc_feature    30410. .30884
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH339524"
     misc_feature    complement(56153. .56817)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH339528"
     misc_feature    57750. .59243
                     /note="wgs_end_extension
                     clone_end:Sp6"
BASE COUNT    68681 a  48371 c  48052 g  67696 t  42831 others
ORIGIN

  Query Match           27.6%;  Score 138.2;  DB 2;  Length 275631;
  Best Local Similarity   64.0%;  Pred. No. 1.3e-23;
  Matches  309;  Conservative    0;  Mismatches  158;  Indels   16;  Gaps    6;

Qy         12 ATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCTAAGTAAGCTCC 71
Db     142970 AGAATATATTTCTGACTCAAAGAGAAATGTGTATTGAGAGCCAGACGCAAAATAAGATAA 142911

Qy         72 AGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTACA 131
Db     142910 AGATCCTCTGCTGGTTTCTTTTTTTCTGGGAGGCAAGTTCTTT-AGACATTGATTTTACA 142852

Qy        132 GGCATATGAATAAAAACTATATTTTGCAG--CATTGTACACTTTTTTTCCTTTTCTAGAA 189
Db     142851 AGCCAACGAATAAGTCTCATATTTTTCAGAACCATGGATTCTATAGATCCTCTCCTACAA 142792

Qy        190 ATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATATCCCTACTTT 249
Db     142791 ATTCAGAACTTCACACATTGGTAGAAATGTTGAGCAAA-------ACGTCCTTTTCTGTT 142739

Qy        250 TCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCTTTTCTTCCT 309
Db     142738 GCGAACGGCTTCCCTCCGCTGGGTCCCTTTGCATCCCCCATGCGTCAGTAGTTTCTTCCC 142679

Qy        310 CATCTCTTTCAGGGGCTGGAGAGGCAGAGGG---AGACAGAGGAGCTGGTACTGCAGAGC 366
Db     142678 C--CTTTCTAAAGGGCTTGAGAGGCAGAGCGCCTCAGCAGCAGAGCGGATCCTATAGCCC 142621

Qy        367 GGTCGTCTGATTGGCTGGACGG-TCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTA 425
Db     142620 TGCCATCTGATTGGCTGCCGGGCGCGTGGCTGGGCTATAAAAGAGACCCCTACAGGCTTG 142561

Qy        426 GCAGGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCT 485
Db     142560 AGAGGCAGAGGCTCAGAGGATTCTGACAGCTTCTTTGCAGAGCAGAAGCAACTGTGCTCA 142501

Qy        486 CAA 488
Db     142500 AAA 142498


RESULT 6
AK093959
LOCUS       AK093959               2599 bp    mRNA    linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ36640 fis, clone TRACH2019151, moderately
            similar to REGULATOR OF G-PROTEIN SIGNALING 4.
ACCESSION   AK093959
VERSION     AK093959.1  GI:21752924
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Suzuki,O., Sasaki,N., Aotsuka,S., Shoji,T., Ichihara,T.,
            Shiohata,N., Matsumoto,K., Hirano,M., Sano,S., Nomura,R.,
            Yoshikawa,Y., Matsumura,Y., Moriya,S., Chiba,E., Momiyama,H.,
            Onogawa,S., Kaeriyama,S., Satoh,N., Matsunawa,H., Takahashi,E.,
            Kataoka,R., Kuga,N., Kuroda,A., Satoh,I., Kamata,K., Takami,S.,
            Terashima,Y., Watanabe,M., Sugiyama,T., Irie,R., Otsuki,T.,
            Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2599)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1. .2599
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="TRACH2019151"
                     /tissue_type="trachea"
                     /clone_lib="TRACH2"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      719 a    520 c    541 g    819 t
ORIGIN

  Query Match           13.9%;  Score 69.4;  DB 9;  Length 2599;
  Best Local Similarity   98.6%;  Pred. No. 1.3e-06;
  Matches   70;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        431 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 490
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCGAAGTACGCTCAAAG 60

Qy        491 CCGAAGCCACA 501
Db         61 CCGAAGCCACA 71


RESULT 7
AK096204
LOCUS       AK096204               3769 bp    mRNA    linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ38885 fis, clone MESAN2017417, moderately
            similar to REGULATOR OF G-PROTEIN SIGNALING 4.
ACCESSION   AK096204
VERSION     AK096204.1  GI:21755635
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M., Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3769)
AUTHORS Isogai,T. and Yamamoto,J.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.
FEATURES Location/Qualifiers
source 1. .3769
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MESAN2017417"
/cell_type="normal mesangial cells (NHMC56046-2)"
/clone_lib="MESAN2"
/note="cloning vector: pME18SFL3~primary culture, normal mesangial cells"
BASE COUNT 1080 a 766 c 731 g 1192 t
ORIGIN

Query Match 13.8%; Score 69; DB 9; Length 3769;
Best Local Similarity 100.0%; Pred. No. 1.7e-06;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        433 GACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCC 492
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCC 60

Qy        493 GAAGCCACA 501
              |||||||||
Db         61 GAAGCCACA 69
```

RESULT 8
AC111539/c
LOCUS AC111539 234560 bp DNA linear HTG 13-MAY-2003
DEFINITION Rattus norvegicus clone CH230-158A12, WORKING DRAFT SEQUENCE, 3 unordered pieces.
ACCESSION AC111539
VERSION AC111539.5 GI:30579258
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 234560)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 234560)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 234560)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 13, 2003 this sequence version replaced gi:25007559.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM

```
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
           -------------- Project Information
            Center project name: GMNF
            Center clone name: CH230-158A12
           -------------- Summary Statistics
            Assembly program: Atlas 3.0;
            Consensus quality: 221222 bases at least Q40
            Consensus quality: 224241 bases at least Q30
            Consensus quality: 226292 bases at least Q20
            Estimated insert size: 231347; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
           --------------
           * NOTE: Estimated insert size may differ from sequence length
           *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 3 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *       1   231343: contig of 231343 bp in length
           *  231344   231443: gap of unknown length
           *  231444   232446: contig of 1003 bp in length
           *  232447   232546: gap of unknown length
           *  232547   234560: contig of 2014 bp in length.
FEATURES             Location/Qualifiers
     source          1. .234560
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-158A12"
     misc_feature    1. .1537
                     /note="wgs_contig"
     misc_feature    3731. .7474
                     /note="wgs_contig"
     misc_feature    225356. .228724
                     /note="wgs_contig"
     misc_feature    230080. .231343
                     /note="wgs_contig"
BASE COUNT    65515 a  49188 c  46995 g  66014 t   6848 others
ORIGIN

  Query Match             11.8%;  Score 59;  DB 2;  Length 234560;
  Best Local Similarity   74.7%;  Pred. No. 0.0006;
  Matches   74;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy        390 CGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGACGCTCAGAGGATTCT 449
              ||| |||||||||||||||||| |||||||||    ||| ||| ||||||||||||||
Db       1533 CGTGGCTGGGCTATAAAAGAGACCCTTACAGGCTTGAGAGGCAGAGGCTCAGAGGATTCT 1474

Qy        450 GACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 488
              ||||   ||||| | |   |  ||  |  || | |  ||
Db       1473 GACAGCTTCTTTGCAGAGCAGAAGCAACTGTGCTCAAAA 1435


RESULT 9
AC100845
LOCUS       AC100845              174328 bp    DNA     linear   PRI 25-MAR-2002
DEFINITION  Homo sapiens chromosome 18, clone RP11-680N20, complete sequence.
ACCESSION   AC100845
VERSION     AC100845.2  GI:19703200
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174328)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-680N20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174328)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 174328)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 25, 2002 this sequence version replaced gi:17048215.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L21577
                Center clone name: 680_N_20
            --------------.
FEATURES             Location/Qualifiers
     source          1. .174328
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-680N20"
```

```
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   102. .187
                     /rpt_family="(TA)n"
     repeat_region   256. .285
                     /rpt_family="AT_rich"
     repeat_region   656. .687
                     /rpt_family="AT_rich"
     repeat_region   complement(865. .960)
                     /rpt_family="MIR"
     repeat_region   complement(971. .1278)
                     /rpt_family="AluY"
     repeat_region   complement(1728. .2029)
                     /rpt_family="AluSg"
     repeat_region   2425. .2653
                     /rpt_family="L1M4"
     repeat_region   3273. .3514
                     /rpt_family="MIR"
     repeat_region   3780. .3833
                     /rpt_family="AT_rich"
     repeat_region   complement(3834. .3939)
                     /rpt_family="L1PA5"
     repeat_region   complement(4079. .4371)
                     /rpt_family="AluJb"
     repeat_region   complement(5881. .6125)
                     /rpt_family="MLT1B"
     repeat_region   complement(6126. .9075)
                     /rpt_family="L1PA13"
     repeat_region   complement(9076. .9252)
                     /rpt_family="MLT1B"
     repeat_region   10440. .10582
                     /rpt_family="HAL1b"
     repeat_region   complement(10583. .10875)
                     /rpt_family="AluSq"
     repeat_region   10876. .10966
                     /rpt_family="HAL1b"
     repeat_region   complement(10995. .11175)
                     /rpt_family="L1MA1"
     repeat_region   11375. .12064
                     /rpt_family="L2"
     repeat_region   12273. .12344
                     /rpt_family="L1ME3A"
     repeat_region   12350. .12394
                     /rpt_family="AT_rich"
     repeat_region   12395. .12876
                     /rpt_family="L2"
     repeat_region   12996. .13055
                     /rpt_family="(TAGA)n"
     repeat_region   13065. .13458
                     /rpt_family="L1ME4A"
     repeat_region   13654. .13760
                     /rpt_family="AluSq/x"
     repeat_region   13869. .13908
                     /rpt_family="AT_rich"
     repeat_region   14076. .14144
                     /rpt_family="MIR"
     repeat_region   14333. .14353
                     /rpt_family="AT_rich"
     repeat_region   complement(14675. .15120)
                     /rpt_family="L3"
     repeat_region   complement(15749. .18697)
                     /rpt_family="L1MA4"
     repeat_region   complement(18698. .18997)
                     /rpt_family="AluSp"
     repeat_region   complement(18998. .20399)
                     /rpt_family="L1MA4"
     repeat_region   20400. .20707
                     /rpt_family="AluY"
     repeat_region   complement(20708. .21184)
                     /rpt_family="L1MA4"
     repeat_region   complement(21178. .22061)
                     /rpt_family="L1MA6"
     repeat_region   complement(22062. .24224)
                     /rpt_family="L1MA4"
     unsure          25698. .25704
                     /note="<30 qual SNGL region"
     repeat_region   25996. .26034
                     /rpt_family="AT_rich"
     repeat_region   complement(26598. .26992)
                     /rpt_family="L2"
     repeat_region   complement(27131. .27273)
                     /rpt_family="L2"
     repeat_region   27307. .27467
                     /rpt_family="L2"
     repeat_region   complement(28167. .28652)
                     /rpt_family="L1MB6"
     repeat_region   28711. .28766
                     /rpt_family="MER112"
     repeat_region   28767. .29060
                     /rpt_family="AluSq"
     repeat_region   29061. .29130
                     /rpt_family="MER112"
     repeat_region   complement(29175. .29326)
                     /rpt_family="MIR"
     repeat_region   complement(29534. .29652)
                     /rpt_family="L1ME4A"
     repeat_region   30421. .30448
                     /rpt_family="AT_rich"
     repeat_region   30459. .30752
                     /rpt_family="AluSx"
     repeat_region   30932. .31224
                     /rpt_family="AluJo"
     repeat_region   31264. .32303
                     /rpt_family="L1PA10"
     repeat_region   33754. .33774
                     /rpt_family="AT_rich"
     repeat_region   35349. .35651
                     /rpt_family="AluSx"
     repeat_region   complement(36456. .36759)
                     /rpt_family="AluSx"
     repeat_region   complement(38255. .38613)
                     /rpt_family="MLT1C"
     repeat_region   complement(39708. .39864)
                     /rpt_family="L1P3"
     repeat_region   40206. .40276

  Query Match             9.6%;  Score 48;  DB 9;  Length 174328;
  Best Local Similarity   60.9%;  Pred. No. 0.32;
  Matches   78;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy        106 AACTTCCTTGATATTTTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATTG 165
              | |||| | | | |||    | ||     |||| | | | || |||||||  |  | |||
Db      44705 AGCTTCTTGGGTTTTTCACATATAGTCTTATATAATTTACAAGTATATTTAACCTCTTTG 44764

Qy        166 TACACTTTTTTTCCTTTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTA 225
              |  || | ||| ||||| | |  ||||||  | ||||||  |||| |||     ||   |
Db      44765 TTGACATGTTTACCTTTCCAATCAATTCTTCAACTCTGATATTGTTGGTTTTCTTATCTT 44824

Qy        226 CATTTTTT 233
                ||||| |
Db      44825 GATTTTAT 44832


RESULT 10
I66494
LOCUS       I66494                  7218 bp    DNA     linear   PAT 28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
```

```
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN

  Query Match             9.5%;  Score 47.6;  DB 6;  Length 7218;
  Best Local Similarity   3.1%;  Pred. No. 0.35;
  Matches    8;  Conservative  158;  Mismatches   92;  Indels    0;  Gaps    0;

Qy         62 AGTAAGCTCCAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTT 121
              ||  ||||   || ::::: ::  : ::::::::::::      :  :::::::  : :::
Db       1053 AGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1112

Qy        122 TTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCTT 181
              ::::::: :   : : :   :     :: : :::: :  : :: : : ::::::::::::
Db       1113 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1172

Qy        182 TTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATATC 241
              ::::     ::::   :::::  : ::  :     : ::   : : :::::::: : ::
Db       1173 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1232

Qy        242 CCTACTTTTCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCTT 301
              ::: ::::::      :::::::: ::: ::: :::  : :: ::  : : ::  ::::
Db       1233 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1292

Qy        302 TTCTTCCTCATCTCTTTC 319
              ::::::::: ::::::::
Db       1293 YYYYYYYYYYYYYYYYYY 1310


RESULT 11
AC095234
LOCUS       AC095234               307979 bp    DNA     linear   HTG 26-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-10H10, *** SEQUENCING IN PROGRESS
            ***, 9 unordered pieces.
ACCESSION   AC095234
VERSION     AC095234.4  GI:22772995
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 307979)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 307979)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 307979)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 10, 2002 this sequence version replaced gi:21722865.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequening reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/).  As a result, the
            sequence may extend beyond the ends of the clone and there may be
            contigs that consist entirely of whole genome shotgun sequence
            reads.  Both end sequences and whole genome shotgun sequence only
            contigs will be indicated in the feature table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDLY
                Center clone name: CH230-10H10
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 216015 bases at least Q40
                Consensus quality: 221740 bases at least Q30
                Consensus quality: 224475 bases at least Q20
                Estimated insert size: 305441; sum-of-contigs estimation
                Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     5706: contig of 5706 bp in length
            *    5707     5806: gap of unknown length
            *    5807   231415: contig of 225609 bp in length
```

```
            *   231416   231515: gap of unknown length
            *   231516   237693: contig of 6178 bp in length
            *   237694   237793: gap of unknown length
            *   237794   287892: contig of 50099 bp in length
            *   287893   287992: gap of unknown length
            *   287993   297230: contig of 9238 bp in length
            *   297231   297330: gap of unknown length
            *   297331   301900: contig of 4570 bp in length
            *   301901   302000: gap of unknown length
            *   302001   303486: contig of 1486 bp in length
            *   303487   303586: gap of unknown length
            *   303587   304819: contig of 1233 bp in length
            *   304820   304919: gap of unknown length
            *   304920   307979: contig of 3060 bp in length.
FEATURES             Location/Qualifiers
     source          1. .307979
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-10H10"
     misc_feature    3381. .5706
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    10057. .10967
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH305632"
     misc_feature    21440. .23858
                     /note="wgs_contig"
     misc_feature    30185. .33987
                     /note="wgs_contig"
     misc_feature    51818. .53151
                     /note="wgs_contig"
     misc_feature    80233. .82290
                     /note="wgs_contig"
     misc_feature    187059. .187447
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH305630"
     misc_feature    231516. .233344
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    234226. .235378
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    236591. .237693
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    298726. .301900
                     /note="wgs_end_extension
                     clone_end:T7"
BASE COUNT    67479 a  47083 c  45283 g  66098 t  82036 others
ORIGIN

  Query Match             9.1%;  Score 45.6;  DB 2;  Length 307979;
  Best Local Similarity   50.4%;  Pred. No. 1.3;
  Matches  138;  Conservative    0;  Mismatches  134;  Indels    2;  Gaps    1;

Qy         10 TAATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCTAAGTAAGCT 69
              | || ||  ||  ||   |  |||    |||||   |  | |||  | | |  || | ||
Db     132895 TTATTTGTGGCCACTTGCCTGAAAATCAAAAGGTGGTCAGTGTCTTCACTTTTGTTAACT 132954

Qy         70 CCAGAATTCCTGCT--GGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTT 127
              ||     |||  ||  ||  ||   | ||  ||  || |    | |||||| || ||
Db     132955 CCCCTTCTCCATCTCAGGACCTAAACATTATAGCTAGAACTGGCATTGATATCTTCTTGG 133014

Qy        128 TACAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAG 187
              | |    || | |||    |  || ||||     ||      ||||  || | |||| |
Db     133015 TTCCTATATTTAATTTTTAAAATTTTTTATTAGATATATTTCTTTACTTACATTTCAAA 133074

Qy        188 AAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATATCCCTACT 247
              |||  |||| |||| ||||
Db     133075 CATTATTCCCCTTCCCAATTTCCTGTATATAAGCACCCATTCCCTTCCCATCCCCCTCCC 133134

Qy        248 TTTCAGAAGGATTTTCTCTGCTCGTTCACTTAAC 281
              |  ||| | |||  | ||   | | || || ||
Db     133135 CTATAGAGGTATTCCCCCTATACATCCAGTTTAC 133168


RESULT 12
AX347195
LOCUS       AX347195               10020 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 2266 from Patent WO0200928.
ACCESSION   AX347195
VERSION     AX347195.1  GI:18495083
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 2266 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .10020
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     3111 a     92 c   2113 g   4704 t
ORIGIN

  Query Match             9.1%;  Score 45.4;  DB 6;  Length 10020;
  Best Local Similarity   51.0%;  Pred. No. 1.3;
  Matches  132;  Conservative    0;  Mismatches  126;  Indels    1;  Gaps    1;

Qy         90 TTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTTACAGGCATATGAATAAAAACT 149
              |||  ||   ||| | |  |   |  | ||||||||||  ||  ||||  ||||  | |
Db       1329 TTTGTTTTAGGGATGAATTTAGTGTTCGAGTTTTTTTTTTTAGTGATATATATAAGGAAT 1388

Qy        150 ATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAGAAATTCTAAACCTCTGACATTG 209
              | | |||    ||         ||||||   ||| |   || | ||||  | ||   |||
Db       1389 TTTTATTGGTTGATATGTTTAGTTTTTTTTATTTTTGTGAAATATAAATTTTTGTTTTTG 1448

Qy        210 GTGGAGACATTGAGTACATTTTTTCCCATATCCCTACTTTTCAGAAGGATTTTCTCTGCT 269
              |  | | |  |  |||   ||||   || |   |  ||||  |      |||| |    |
Db       1449 ATTTA-AGAAAGGTTACGCGTTTTTTTATTTTTTTGTTTTTTTGTTTTTTTTTTTTATTT 1507

Qy        270 CGTTCACTTAACATTGCTGATGCGTCAGTCTTTTCTTCCTCATCTCTTTCAGGGGCTGGA 329
              |   || |  ||  |  |   |  | |||| |  | | |  ||| |     ||||
Db       1508 TTATTTTTTATTTTTTTTTTTTTTTATTTTTTATAATAAACGTTTTAAAATTATGGA 1567

Qy        330 GAGGCAGAGGGAGACAGAG 348
              |||  |||| |||| ||||
Db       1568 GAGAGAGAGAGAGAGAGAG 1586


RESULT 13
AC116367/c
LOCUS       AC116367               134971 bp    DNA     linear   HTG 31-OCT-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 11 clone
            OSJNBa0059H21, *** SEQUENCING IN PROGRESS ***, 5 ordered pieces.
ACCESSION   AC116367
VERSION     AC116367.9  GI:24431629
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 134971)
  AUTHORS   Buell,C.R., Yuan,Q., Ouyang,S., Liu,J., Gansberger,K., Jones,K.M.,
            Overton II,L., Tsitrin,T., Kim,M., Bera,J., Jin,S., Fadrosh,D.W.,
```

```
            Tallon,L., Koo,H., Zismann,V., Hsiao,J., Blunt,S., Vanaken,S.,
            Riedmuller,S.B., Utterbach,T., Feldblyum,T., Yang,Q., Haas,B.,
            Suh,B., Peterson,J., Quackenbush,J., White,O., Salzberg,S. and
            Fraser,C.
  TITLE     Oryza sativa ssp. japonica cv. Nipponbare OSJNBa0059H21 BAC genomic
            sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 134971)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 134971)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On Oct 31, 2002 this sequence version replaced gi:23308041.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    29813: contig of 29813 bp in length
            *    29814    29913: gap of unknown length
            *    29914    53824: contig of 23911 bp in length
            *    53825    53924: gap of unknown length
            *    53925    71255: contig of 17331 bp in length
            *    71256    71355: gap of unknown length
            *    71356   115072: contig of 43717 bp in length
            *   115073   115172: gap of unknown length
            *   115173   134971: contig of 19799 bp in length.
FEATURES             Location/Qualifiers
     source          1. .134971
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="11"
                     /clone="OSJNBa0059H21"
                     /note="japonica cultivar-group"
BASE COUNT    37377 a  28943 c  29426 g  38166 t   1059 others
ORIGIN

  Query Match             9.1%;  Score 45.4;  DB 2;  Length 134971;
  Best Local Similarity   25.7%;  Pred. No. 1.4;
  Matches   61;  Conservative   75;  Mismatches  101;  Indels    0;  Gaps    0;

Qy         83 TGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTTACAGGCATATGAAT 142
              |:: | :::|  ||  |::  :::: :|  |: |  |    ||   || |    | |||
Db      29069 TKRAAAWWWTKKTTKRARKTKKYMWWWTTTTWAAATTCAAATTCAAACTGTTCAAAAAAT 29010

Qy        143 AAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAGAAATTCTAAACCTCT 202
                |  || |      | ||   || | || |     ||  :::::::: :||:: | |
Db      29009 GTTATATAGAAAAATGACCAAAATAAAAGTTATAAATCTTRWKRRRWWWWWWAAMYTTTT 28950

Qy        203 GACATTGGTGGAGACATTGAGTACATTTTTTCCCATATCCCTACTTTTCAGAAGGATTTT 262
                  :    |:::::::: :::||: :  ::::    | : :|::| |:  ::|||
Db      28949 TTTGMMAMYTTTTYCWWWKRRKTYMWTTWRKWWYMWWAAAWTYMAWTYMAAAWTTWWTTT 28890

Qy        263 CTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCTTTTCTTCCTCATCTCTTTC 319
              ::  ::: ::| ::::: |:| :   : : |:: | |||:| ::| : ::::|||:
Db      28889 TKSCCYYYYYYCYYYYMWWAWTTYCCYYCMWTYMWTTTTTYCCYYCMMYYYYYTTTY 28833


RESULT 14
AC022764
LOCUS       AC022764              198464 bp    DNA     linear   HTG 30-OCT-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-310M15 map 17, WORKING DRAFT
            SEQUENCE.
ACCESSION   AC022764
VERSION     AC022764.4  GI:11038547
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 198464)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-310M15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 198464)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 30, 2000 this sequence version replaced gi:7249221.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5848
                Center clone name: 310_M_15
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 32% of reads
                Sequencing vector: Plasmid; n/a; 68% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 198356 bases at least Q40
                Consensus quality: 198459 bases at least Q30
                Consensus quality: 198464 bases at least Q20
                Insert size: 210000; agarose-fp
                Insert size: 198464; sum-of-contigs
                Quality coverage: 9.2 in Q20 bases; agarose-fp
                Quality coverage: 9.7 in Q20 b.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1   198464: contig of 198464 bp in length.
            assembly_fragment.
FEATURES             Location/Qualifiers
     source          1. .198464
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
```

```
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-310M15"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    65747 a  36134 c  36064 g  60519 t
ORIGIN

  Query Match             9.0%;  Score 45.2;  DB 2;  Length 198464;
  Best Local Similarity   50.0%;  Pred. No. 1.6;
  Matches  113;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy          60 TAAGTAAGCTCCAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATAT 119
               | | | |  || |  ||   |  | |  |||  ||||| |   |||  |   |  ||||
Db       55513 TCAATCACTTCTACGATCTGTAATAATTTTTTGTCTTCCTGAGTGCATATATATATATAT 55572

Qy         120 TTTTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCC 179
                ||||||||  |   | |||  ||||     || ||||    || | || | ||| |||
Db       55573 ATTTTTTTTCCTTTCGTATATATAATTGTCATTTTTTATTTCAATTTAAAATTTATTTGT 55632

Qy         180 TTTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATA 239
                |  || | |  |  ||  | | | | |  |  | | | |   ||  | |||| || |
Db       55633 ATGACTTGTACATAAAATTTTATTAGAATTTTTAATATACTCTATAAGTATTTTTCCCTC 55692

Qy         240 TCCCTACTTTTCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTG 285
               |  ||||  |   ||   | |  | ||    ||    | | ||
Db       55693 TGGTTACTAATGTTAATTTTATCTTTTGTCTATTTTTCTTAAATGG 55738


RESULT 15
AL451076/c
LOCUS       AL451076               203581 bp    DNA     linear   ROD 29-JUN-2002
DEFINITION  Mouse DNA sequence from clone RP23-43O20 on chromosome X, complete
            sequence.
ACCESSION   AL451076
VERSION     AL451076.14  GI:20068432
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 203581)
  AUTHORS   Phillimore,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 7, 2002 this sequence version replaced gi:14161186.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-43O20 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: UK Medical Research Council
            Center code: UK-MRC
            Web site: http://mrcseq.har.mrc.ac.uk
            Contact: mouseq@har.mrc.ac.uk
            --------------.
FEATURES             Location/Qualifiers
     source          1. .203581
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="X"
                     /clone="RP23-43O20"
                     /clone_lib="RPCI-23"
BASE COUNT    58942 a  42735 c  44981 g  56923 t
ORIGIN

  Query Match             9.0%;  Score 45.2;  DB 10;  Length 203581;
  Best Local Similarity   54.9%;  Pred. No. 1.6;
  Matches   89;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy         105 CAACTTCCTTGATATTTTTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATT 164
               |||||||||   | |||   ||       | |    | |    | | | |||| ||||||
Db       40021 CAACTTCCTGTCTTTTTAGCTTGCCTCAGAACTCTACTTCTGAATTTTTTTTTCAGCATA 39962

Qy         165 GTACACTTTTTTTCCTTTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGT 224
               ||  ||||||||  ||| ||  ||| |       | |||  ||  ||   ||||   ||
Db       39961 TTAACCTTTTTTTTTTTTGCTTTAAAATACTTTATTGTGATCTTTATGATAACATCAGGT 39902

Qy         225 ACATTTTTTCCCATATCCCTACTTTTCAGAAGGATTTTCTCT 266
               |  | |||    ||||   |  |||  |||  |  |||||||
Db       39901 ATTTATTTATTTATATATGTTTTTTCAAGACAGGGTTTCTCT 39860


Search completed: November  7, 2003, 11:11:19
Job time : 1969.4 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 01:54:16 ; Search time 1962.4 Seconds
(without alignments)
10444.222 Million cell updates/sec

Title: US-09-939-209A-3_COPY_15000_15500
Perfect score: 501
Sequence: 1 tggcagagaactctctggat.....tcaacctgatggagaaggat 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
	1	501	100.0	20300	6	AX451337	AX451337 Sequence
	2	501	100.0	165329	9	AL583850	AL583850 Human DNA
c	3	501	100.0	191699	2	AC031977	AC031977 Homo sapi
	4	499.4	99.7	3769	9	AK096204	AK096204 Homo sapi
c	5	145.4	29.0	207387	2	AC115766	AC115766 Mus muscu
c	6	130.6	26.1	275631	2	AC125563	AC125563 Rattus no
	7	113	22.6	618	9	AF493928	AF493928 Homo sapi
	8	113	22.6	618	9	BT007025	BT007025 Homo sapi
	9	113	22.6	618	12	BT007756	BT007756 Synthetic
	10	113	22.6	800	6	AR270528	AR270528 Sequence
	11	113	22.6	800	9	HSU27768	U27768 Human RGP4
	12	113	22.6	1869	9	BC000737	BC000737 Homo sapi
	13	113	22.6	2753	9	BC051869	BC051869 Homo sapi
	14	113	22.6	2934	6	AX451335	AX451335 Sequence
	15	111.8	22.3	2599	9	AK093959	AK093959 Homo sapi
	16	100.2	20.0	1489	10	RNU27767	U27767 Rattus norv
	17	100.2	20.0	2919	10	AF117211	AF117211 Rattus no
	18	98.6	19.7	2781	10	BC003882	BC003882 Mus muscu
	19	97	19.4	630	10	AB004315	AB004315 Mus muscu
	20	89.8	17.9	201	10	RNU32327	U32327 Rattus norv
	21	78.4	15.6	150	5	AF090091	AF090091 Gallus ga
	22	76.8	15.3	254	5	AF090081	AF090081 Gallus ga
	23	65	13.0	964	5	AF263451	AF263451 Xenopus l
	24	54.2	10.8	275631	2	AC125563	AC125563 Rattus no
	25	53.4	10.7	3653	9	AF009356	AF009356 Homo sapi
c	26	53.4	10.7	102812	2	AL158215	AL158215 Homo sapi
	27	53.4	10.7	186051	2	AC015683	AC015683 Homo sapi
c	28	52.6	10.5	165329	9	AL583850	AL583850 Human DNA
	29	52.6	10.5	191699	2	AC031977	AC031977 Homo sapi
	30	52.6	10.5	192394	2	BX119907	BX119907 Danio rer
	31	52.2	10.4	2078	4	SSC549925	AJ549925 Sus scrof
c	32	51.8	10.3	147047	9	AL353778	AL353778 Human DNA
	33	51.6	10.3	6905	10	AY138504	AY138504 Mus muscu
c	34	51.6	10.3	50095	2	AC074333	AC074333 Mus muscu
c	35	51.6	10.3	132762	10	AL844529	AL844529 Mouse DNA
c	36	50.6	10.1	144798	2	AC141948	AC141948 Rattus no
c	37	50.6	10.1	188016	2	AC121256	AC121256 Mus muscu
c	38	50.6	10.1	246357	2	AC109089	AC109089 Rattus no
	39	49.8	9.9	201	5	AB038435	AB038435 Xenopus l
	40	49.6	9.9	3494	10	AB006013	AB006013 Rattus no
	41	49	9.8	378	6	BD168779	BD168779 New disea
	42	49	9.8	543	6	BD16877B	BD168778 New disea
	43	49	9.8	546	10	AF241259	AF241259 Rattus no
	44	49	9.8	817	6	BD168782	BD168782 New disea
c	45	48.4	9.7	106777	2	AC136241	AC136241 Rattus no

ALIGNMENTS

RESULT 1
AX451337
LOCUS AX451337 20300 bp DNA linear PAT 03-JUL-2002
DEFINITION Sequence 3 from Patent WO0216653.
ACCESSION AX451337
VERSION AX451337.1 GI:21698388
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Levitt,P.R., Mirnics,K., Kodavali,V.C. and Nimgaonkar,V.L.
TITLE Methods and systems for facilitating the diagnosis and treatment of schizophrenia
JOURNAL Patent: WO 0216653-A 3 28-FEB-2002;

```
            University of Pittsburgh (US)
FEATURES             Location/Qualifiers
     source          1. .20300
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="A genomic sequence containing RGS4 nucleic acid
                     sequence and sequ ences upstream and downstream to the
                     RGS4 nucleic acid sequence"
BASE COUNT     6157 a   4102 c   3775 g   6266 t
ORIGIN

  Query Match           100.0%;  Score 501;  DB 6;  Length 20300;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-138;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGGCAGAGAACTCTCTGGATCTTAGTGAAGGTTCCTAGAATAGTGGAGCTGACTATCATA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15000 TGGCAGAGAACTCTCTGGATCTTAGTGAAGGTTCCTAGAATAGTGGAGCTGACTATCATA 15059

Qy         61 ATCTTGACAACCCCAAATAAATCAGTTTTTTAAAAAATCTCTTTTATCCATGTGGCTTAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15060 ATCTTGACAACCCCAAATAAATCAGTTTTTTAAAAAATCTCTTTTATCCATGTGGCTTAC 15119

Qy        121 CATAACCTCCCTGCATGAATTTTTCTGATGAATCTCCCCAATTTGTTAGACAGAACAGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15120 CATAACCTCCCTGCATGAATTTTTCTGATGAATCTCCCCAATTTGTTAGACAGAACAGAA 15179

Qy        181 GATCTTGCCCTGCTCTCTCTAAAGCAGAAAGGTTCATTCTGAACCTTTCATACTCTCTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15180 GATCTTGCCCTGCTCTCTCTAAAGCAGAAAGGTTCATTCTGAACCTTTCATACTCTCTCA 15239

Qy        241 CATGTGCCAAGGAGGACCCCAATGTCACTTTTGTTTTTTGCTTCTGAAATACAGAGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15240 CATGTGCCAAGGAGGACCCCAATGTCACTTTTGTTTTTTGCTTCTGAAATACAGAGGGTG 15299

Qy        301 CACTGCCACTTACAAGTCACTACAAAGCATACAGGCTTGCATCCTCAACAGGGATATAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15300 CACTGCCACTTACAAGTCACTACAAAGCATACAGGCTTGCATCCTCAACAGGGATATAGG 15359

Qy        361 TCTAATGAAGCCTTGGCCTTTGCCCCTCAGGTGAACCTGGATTCTTGCACCAGGGAAGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15360 TCTAATGAAGCCTTGGCCTTTGCCCCTCAGGTGAACCTGGATTCTTGCACCAGGGAAGAG 15419

Qy        421 ACAAGCCGGAACATGCTAGAGCCTACAATAACCTGCTTTGATGAGGCCCAGAAGAAGATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15420 ACAAGCCGGAACATGCTAGAGCCTACAATAACCTGCTTTGATGAGGCCCAGAAGAAGATT 15479

Qy        481 TTCAACCTGATGGAGAAGGAT 501
              |||||||||||||||||||||
Db      15480 TTCAACCTGATGGAGAAGGAT 15500


RESULT 2
AL583850
LOCUS       AL583850              165329 bp    DNA     linear   PRI 15-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-430G6 on chromosome 1, complete
            sequence.
ACCESSION   AL583850
VERSION     AL583850.5  GI:16973044
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165329)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 16, 2001 this sequence version replaced gi:15020514.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-430G6 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-430G6 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP11-430G6 is at 165329 in this
            sequence. The true right end of clone RP11-331H2 is at 2000 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .165329
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-430G6"
                     /clone_lib="RPCI-11.2"
BASE COUNT    51144 a  30897 c  31439 g  51849 t
ORIGIN

  Query Match           100.0%;  Score 501;  DB 9;  Length 165329;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-138;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGGCAGAGAACTCTCTGGATCTTAGTGAAGGTTCCTAGAATAGTGGAGCTGACTATCATA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86681 TGGCAGAGAACTCTCTGGATCTTAGTGAAGGTTCCTAGAATAGTGGAGCTGACTATCATA 86740

Qy         61 ATCTTGACAACCCCAAATAAATCAGTTTTTTAAAAAATCTCTTTTATCCATGTGGCTTAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86741 ATCTTGACAACCCCAAATAAATCAGTTTTTTAAAAAATCTCTTTTATCCATGTGGCTTAC 86800

Qy        121 CATAACCTCCCTGCATGAATTTTTCTGATGAATCTCCCCAATTTGTTAGACAGAACAGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86801 CATAACCTCCCTGCATGAATTTTTCTGATGAATCTCCCCAATTTGTTAGACAGAACAGAA 86860

Qy        181 GATCTTGCCCTGCTCTCTCTAAAGCAGAAAGGTTCATTCTGAACCTTTCATACTCTCTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86861 GATCTTGCCCTGCTCTCTCTAAAGCAGAAAGGTTCATTCTGAACCTTTCATACTCTCTCA 86920

Qy        241 CATGTGCCAAGGAGGACCCCAATGTCACTTTTGTTTTTTGCTTCTGAAATACAGAGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86921 CATGTGCCAAGGAGGACCCCAATGTCACTTTTGTTTTTTGCTTCTGAAATACAGAGGGTG 86980

Qy        301 CACTGCCACTTACAAGTCACTACAAAGCATACAGGCTTGCATCCTCAACAGGGATATAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86981 CACTGCCACTTACAAGTCACTACAAAGCATACAGGCTTGCATCCTCAACAGGGATATAGG 87040

Qy        361 TCTAATGAAGCCTTGGCCTTTGCCCCTCAGGTGAACCTGGATTCTTGCACCAGGGAAGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87041 TCTAATGAAGCCTTGGCCTTTGCCCCTCAGGTGAACCTGGATTCTTGCACCAGGGAAGAG 87100
```

```
Qy        421 ACAAGCCGGAACATGCTAGAGCCTACAATAACCTGCTTTGATGAGGCCCAGAAGAAGATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87101 ACAAGCCGGAACATGCTAGAGCCTACAATAACCTGCTTTGATGAGGCCCAGAAGAAGATT 87160

Qy        481 TTCAACCTGATGGAGAAGGAT 501
              |||||||||||||||||||||
Db      87161 TTCAACCTGATGGAGAAGGAT 87181


RESULT 3
AC031977/c
LOCUS       AC031977               191699 bp    DNA     linear   HTG 12-APR-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-288O18, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC031977
VERSION     AC031977.7  GI:13194952
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
            Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
            Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
            Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Mar 4, 2001 this sequence version replaced gi:9665085.
            -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 880
                 Center clone name: RP11-288O18
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513; 100% of reads
                 Sequencing Vector: plasmid; plasmid_accession; 0% of reads
                 Chemistry: Dye-primer; 1% of reads
                 Chemistry: Dye-terminator Big Dye; 99% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 190680 bases at least Q40
                 Consensus quality: 191287 bases at least Q30
                 Consensus quality: 191336 bases at least Q20
                 Insert size: 195548; agarose-fp
                 Insert size: 191499; sum-of-contigs
                 Quality coverage:  7.9x in Q20 bases; agarose-fp
                 Quality coverage:  8.1x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    12646: contig of 12646 bp in length
            *    12647    12746: gap of unknown length
            *    12747    94961: contig of 82215 bp in length
            *    94962    95061: gap of unknown length
            *    95062   191699: contig of 96638 bp in length
FEATURES             Location/Qualifiers
     source          1. .191699
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-288O18"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .12646
                     /note="assembly_name:Contig5"
     misc_feature    12747. .94961
                     /note="assembly_name:Contig6
                     clone_end:T7"
     misc_feature    95062. .191699
                     /note="assembly_name:Contig7
                     clone_end:SP6"
BASE COUNT    57654 a  36385 c  36166 g  61293 t    201 others
ORIGIN

  Query Match             100.0%;  Score 501;  DB 2;  Length 191699;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-138;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGGCAGAGAACTCTCTGGATCTTAGTGAAGGTTCCTAGAATAGTGGAGCTGACTATCATA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189481 TGGCAGAGAACTCTCTGGATCTTAGTGAAGGTTCCTAGAATAGTGGAGCTGACTATCATA 189422

Qy         61 ATCTTGACAACCCCAAATAAATCAGTTTTTTAAAAAATCTCTTTTATCCATGTGGCTTAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189421 ATCTTGACAACCCCAAATAAATCAGTTTTTTAAAAAATCTCTTTTATCCATGTGGCTTAC 189362

Qy        121 CATAACCTCCCTGCATGAATTTTTCTGATGAATCTCCCCAATTTGTTAGACAGAACAGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189361 CATAACCTCCCTGCATGAATTTTTCTGATGAATCTCCCCAATTTGTTAGACAGAACAGAA 189302

Qy        181 GATCTTGCCCTGCTCTCTCTAAAGCAGAAAGGTTCATTCTGAACCTTTCATACTCTCTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189301 GATCTTGCCCTGCTCTCTCTAAAGCAGAAAGGTTCATTCTGAACCTTTCATACTCTCTCA 189242

Qy        241 CATGTGCCAAGGAGGACCCCAATGTCACTTTTGTTTTTTGCTTCTGAAATACAGAGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189241 CATGTGCCAAGGAGGACCCCAATGTCACTTTTGTTTTTTGCTTCTGAAATACAGAGGGTG 189182

Qy        301 CACTGCCACTTACAAGTCACTACAAAGCATACAGGCTTGCATCCTCAACAGGGATATAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189181 CACTGCCACTTACAAGTCACTACAAAGCATACAGGCTTGCATCCTCAACAGGGATATAGG 189122

Qy        361 TCTAATGAAGCCTTGGCCTTTGCCCCTCAGGTGAACCTGGATTCTTGCACCAGGGAAGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189121 TCTAATGAAGCCTTGGCCTTTGCCCCTCAGGTGAACCTGGATTCTTGCACCAGGGAAGAG 189062

Qy        421 ACAAGCCGGAACATGCTAGAGCCTACAATAACCTGCTTTGATGAGGCCCAGAAGAAGATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189061 ACAAGCCGGAACATGCTAGAGCCTACAATAACCTGCTTTGATGAGGCCCAGAAGAAGATT 189002

Qy        481 TTCAACCTGATGGAGAAGGAT 501
              |||||||||||||||||||||
Db     189001 TTCAACCTGATGGAGAAGGAT 188981


RESULT 4
AK096204
LOCUS       AK096204               3769 bp    mRNA    linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ38885 fis, clone MESAN2017417, moderately
            similar to REGULATOR OF G-PROTEIN SIGNALING 4.
ACCESSION   AK096204
VERSION     AK096204.1  GI:21755635
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1
  AUTHORS   Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M.,
            Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K.,
            Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
            Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
            Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
            Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
            Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
            Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3769)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1. .3769
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MESAN2017417"
                     /cell_type="normal mesangial cells (NHMC56046-2)"
                     /clone_lib="MESAN2"
                     /note="cloning vector: pME18SFL3~primary culture, normal
                     mesangial cells"
BASE COUNT     1080 a    766 c    731 g   1192 t
ORIGIN

  Query Match            99.7%;  Score 499.4;  DB 9;  Length 3769;
  Best Local Similarity  99.8%;  Pred. No. 3.7e-138;
  Matches  500;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TGGCAGAGAACTCTCTGGATCTTAGTGAAGGTTCCTAGAATAGTGGAGCTGACTATCATA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1114 TGGCAGAGAACTCTCTGGATCTTAGTGAAGGTTCCTAGAATAGTGGAGCTGACTATCATA 1173

Qy         61 ATCTTGACAACCCCAAATAAATCAGTTTTTTAAAAAATCTCTTTTATCCATGTGGCTTAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1174 ATCTTGACAACCCCAAATAAATCAGTTTTTTAAAAAATCTCTTTTATCCATGTGGCTTAC 1233

Qy        121 CATAACCTCCCTGCATGAATTTTTCTGATGAATCTCCCCAATTTGTTAGACAGAACAGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1234 CATAACCTCCCTGCATGAATTTTTCTGATGAATCTCCCCAATTTGTTAGACAGAACAGAA 1293

Qy        181 GATCTTGCCCTGCTCTCTCTAAAGCAGAAAGGTTCATTCTGAACCTTTCATACTCTCTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1294 GATCTTGCCCTGCTCTCTCTAAAGCAGAAAGGTTCATTCTGAACCTTTCATACTCTCTCA 1353

Qy        241 CATGTGCCAAGGAGGACCCCAATGTCACTTTTGTTTTTTGCTTCTGAAATACAGAGGGTG 300
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       1354 CATGTGCCAAGGAGGACCCCAATGTCACTTTTGTTTTTTGCTTCTCAAATACAGAGGGTG 1413

Qy        301 CACTGCCACTTACAAGTCACTACAAAGCATACAGGCTTGCATCCTCAACAGGGATATAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1414 CACTGCCACTTACAAGTCACTACAAAGCATACAGGCTTGCATCCTCAACAGGGATATAGG 1473

Qy        361 TCTAATGAAGCCTTGGCCTTTGCCCCTCAGGTGAACCTGGATTCTTGCACCAGGGAAGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1474 TCTAATGAAGCCTTGGCCTTTGCCCCTCAGGTGAACCTGGATTCTTGCACCAGGGAAGAG 1533

Qy        421 ACAAGCCGGAACATGCTAGAGCCTACAATAACCTGCTTTGATGAGGCCCAGAAGAAGATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1534 ACAAGCCGGAACATGCTAGAGCCTACAATAACCTGCTTTGATGAGGCCCAGAAGAAGATT 1593

Qy        481 TTCAACCTGATGGAGAAGGAT 501
              |||||||||||||||||||||
Db       1594 TTCAACCTGATGGAGAAGGAT 1614


RESULT 5
AC115766/c
LOCUS       AC115766               207387 bp    DNA     linear   HTG 23-MAR-2003
DEFINITION  Mus musculus clone RP23-82I24, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   AC115766
VERSION     AC115766.4  GI:29164623
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 207387)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-82I24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 207387)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 207387)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 23, 2003 this sequence version replaced gi:28191504.
```

```
        All repeats were identified using RepeatMasker:
        Smit, A.F.A. & Green, P. (1996-1997)
        http://ftp.genome.washington.edu/RM/RepeatMasker.html
        -------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
        -------------- Project Information
            Center project name: L23317
            Center clone name: 82_I_24
        -------------- Summary Statistics
            Sequencing vector: Plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 205497 bases at least Q40
            Consensus quality: 206243 bases at least Q30
            Consensus quality: 206432 bases at least Q20
            Insert size: 205000; agarose-fp
            Insert size: 206587; sum-of-contigs
            Quality coverage: 8.8 in Q20 bases; agarose-fp
            Quality coverage: 8.7 in Q20 bases; sum-of-contigs
        --------------.
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 9 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *        1    61925: contig of 61925 bp in length
        *    61926    62025: gap of 100 bp
        *    62026    64603: contig of 2578 bp in length
        *    64604    64703: gap of 100 bp
        *    64704    68576: contig of 3873 bp in length
        *    68577    68676: gap of 100 bp
        *    68677    71936: contig of 3260 bp in length
        *    71937    72036: gap of 100 bp
        *    72037    76569: contig of 4533 bp in length
        *    76570    76669: gap of 100 bp
        *    76670    82563: contig of 5894 bp in length
        *    82564    82663: gap of 100 bp
        *    82664   119650: contig of 36987 bp in length
        *   119651   119750: gap of 100 bp
        *   119751   165174: contig of 45424 bp in length
        *   165175   165274: gap of 100 bp
        *   165275   207387: contig of 42113 bp in length.
FEATURES             Location/Qualifiers
     source          1. .207387
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-82I24"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .61925
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    62026. .64603
                     /note="assembly_fragment"
     misc_feature    64704. .68576
                     /note="assembly_fragment"
     misc_feature    68677. .71936
                     /note="assembly_fragment"
     misc_feature    72037. .76569
                     /note="assembly_fragment"
     misc_feature    76670. .82563
                     /note="assembly_fragment"
     misc_feature    82664. .119650
                     /note="assembly_fragment"
     misc_feature    119751. .165174
                     /note="assembly_fragment"
     misc_feature    165275. .207387
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    60516 a  42224 c  43121 g  60726 t    800 others
ORIGIN

  Query Match             29.0%;  Score 145.4;  DB 2;  Length 207387;
  Best Local Similarity   74.3%;  Pred. No. 3.7e-32;
  Matches  197;  Conservative    0;  Mismatches   66;  Indels    2;  Gaps    1;

Qy        239 CACATGTGCCAAGGAGGACCCCAATGTCACTTTTGTTTTTTGCTTCTGAAA--TACAGAG 296
              ||||| || || ||  ||||| ||||||  |   | |  ||        ||||  || |||
Db     111101 CACATATGTCAGGGCAGACCCAAATGTCCCGGGTTTCATTCCGGGGAGAAATGTAAAGAA 111042

Qy        297 GGTGCACTGCCACTTACAAGTCACTACAAAGCATACAGGCTTGCATCCTCAACAGGGATA 356
              | | |  |  | | ||   | | |||| | ||| ||    || |||||||||
Db     111041 AGGGTCCCAGCCACTCCTTGTTCATTCCAAGCCTGCAGCCTCCTCTCTTCAACAGGGCAC 110982

Qy        357 TAGGTCTAATGAAGCCTTGGCCTTTGCCCCTCAGGTGAACCTGGATTCTTGCACCAGGGA 416
              ||| ||||||| | |||||| |||||| | ||||||||||||||| |||||||||| ||
Db     110981 TAGATCTAATGCAACCTTGGTCTTTGCACTTCAGGTGAACCTGGACTCTTGCACCAGAGA 110922

Qy        417 AGAGACAAGCCGGAACATGCTAGAGCCTACAATAACCTGCTTTGATGAGGCCCAGAAGAA 476
              |||||||||||||||||| || |||| |||||||||| |||||||| ||||||||||
Db     110921 GGAGACAAGCCGGAACATGTTACAGCCCACAATAACCTGTTTTGATGAAGCCCAGAAGAA 110862

Qy        477 GATTTTCAACCTGATGGAGAAGGAT 501
              ||||||||||||||||||| ||||||
Db     110861 GATTTTCAACCTGATGGAAAAGGAT 110837


RESULT 6
AC125563/c
LOCUS       AC125563               275631 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-9B12, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC125563
VERSION     AC125563.4  GI:24817906
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 275631)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
```

```
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 275631)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 275631)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:23096552.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDBK
                Center clone name: CH230-9B12
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 228152 bases at least Q40
                Consensus quality: 230310 bases at least Q30
                Consensus quality: 231872 bases at least Q20
                Estimated insert size: 235029; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   192878: contig of 192878 bp in length
            *    192879   192978: gap of unknown length
            *    192979   270295: contig of 77317 bp in length
            *    270296   270395: gap of unknown length
            *    270396   272405: contig of 2010 bp in length
            *    272406   272505: gap of unknown length
            *    272506   275631: contig of 3126 bp in length.
FEATURES             Location/Qualifiers
     source          1. .275631
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-9B12"
     misc_feature    30410. .30884
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH339524"
     misc_feature    complement(56153. .56817)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH339528"
     misc_feature    57750. .59243
                     /note="wgs_end_extension
                     clone_end:Sp6"
BASE COUNT    68681 a  48371 c  48052 g  67696 t  42831 others
ORIGIN

  Query Match             26.1%;  Score 130.6;  DB 2;  Length 275631;
  Best Local Similarity   85.8%;  Pred. No. 1e-27;
  Matches  145;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy        333 AGGCTTGCATCCTCAACAGGGATATAGGTCTAATGAAGCCTTGGCCTTTGCCCCTCAGGT 392
              || ||    || ||||||| | | ||| ||||||| | |||||| |||||| | | ||||
Db     138556 AGCCTCCTCTCTTCAACAGAGCTCTAGCTCTAATGCAACCTTGGTCTTTGCACTTTAGGT 138497

Qy        393 GAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAATAAC 452
              |||||||||||||||||||| || |||||||||||||||||| ||||||| || |||||
Db     138496 GAACCTGGATTCTTGCACCAGAGAGGAGACAAGCCGGAACATGTTAGAGCCCACGATAAC 138437

Qy        453 CTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 501
              ||| |||||||| |||||||||||||||||||||||||||||| ||||||
Db     138436 CTGTTTTGATGAAGCCCAGAAGAAGATTTTCAACCTGATGGAAAAGGAT 138388


RESULT 7
AF493928
LOCUS       AF493928                 618 bp    mRNA    linear   PRI 14-APR-2002
DEFINITION  Homo sapiens regulator of G protein signalling 4 (RGS4) mRNA,
            complete cds.
ACCESSION   AF493928
VERSION     AF493928.1  GI:20147748
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 618)
  AUTHORS   Puhl,H.L. III, Ikeda,S.R. and Aronstam,R.S.
  TITLE     Regulator of G protein signalling 4 (RGS4), human
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 618)
  AUTHORS   Puhl,H.L. III, Ikeda,S.R. and Aronstam,R.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2002) cDNA Resource Center, Guthrie Research
            Institute, One Guthrie Square, Sayre, PA 18840, USA
FEATURES             Location/Qualifiers
     source          1. .618
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
```

```
                     /db_xref="taxon:9606"
                     /tissue_type="brain"
     gene            1. .618
                     /gene="RGS4"
     CDS             1. .618
                     /gene="RGS4"
                     /codon_start=1
                     /product="regulator of G protein signalling 4"
                     /protein_id="AAM12642.1"
                     /db_xref="GI:20147749"
                     /translation="MCKGLAGLPASCLRSAKDMKHRLGFLLQKSDSCEHNSSHNKKDK
                     VVICQRVSQEEVKKWAESLENLISHECGLAAFKAFLKSEYSEENIDFWISCEEYKKIK
                     SPSKLSPKAKKIYNEFISVQATKEVNLDSCTREETSRNMLEPTITCFDEAQKKIFNLM
                     EKDSYRRFLKSRFYLDLVNPSSCGAEKQKGAKSSADCASLVPQCA"
BASE COUNT      191 a    133 c    150 g    144 t
ORIGIN

  Query Match             22.6%;  Score 113;  DB 9;  Length 618;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-22;
  Matches  113;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 436

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 501
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 489


RESULT 8
BT007025
LOCUS       BT007025                618 bp    mRNA    linear   PRI 13-MAY-2003
DEFINITION  Homo sapiens regulator of G-protein signalling 4 mRNA, complete
            cds.
ACCESSION   BT007025
VERSION     BT007025.1  GI:30582888
KEYWORDS    FLI_CDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 618)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 618)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT     This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been  directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI  and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone:  'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES             Location/Qualifiers
     source          1. .618
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GH00329X1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
                     /lab_host="DH5alpha T1 resistant"
                     /note="Vector: pDNR-Dual"
     CDS             1. .618
                     /codon_start=1
                     /product="regulator of G-protein signalling 4"
                     /protein_id="AAP35671.1"
                     /db_xref="GI:30582889"
                     /translation="MCKGLAGLPASCLRSAKDMKHRLGFLLQKSDSCEHNSSHNKKDK
                     VVICQRVSQEEVKKWAESLENLISHECGLAAFKAFLKSEYSEENIDFWISCEEYKKIK
                     SPSKLSPKAKKIYNEFISVQATKEVNLDSCTREETSRNMLEPTITCFDEAQKKIFNLM
                     EKDSYRRFLKSRFYLDLVNPSSCGAEKQKGAKSSADCASLVPQCA"
BASE COUNT      190 a    133 c    151 g    144 t
ORIGIN

  Query Match             22.6%;  Score 113;  DB 9;  Length 618;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-22;
  Matches  113;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 436

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 501
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 489


RESULT 9
BT007756
LOCUS       BT007756                618 bp    mRNA    linear   SYN 13-MAY-2003
DEFINITION  Synthetic construct Homo sapiens regulator of G-protein signalling
            4 mRNA, partial cds.
ACCESSION   BT007756
VERSION     BT007756.1  GI:30584350
KEYWORDS    FLI_CDNA.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 618)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 618)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT     This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been  directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI  and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone:  'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES             Location/Qualifiers
     source          1. .618
                     /organism="synthetic construct"
                     /mol_type="mRNA"
                     /db_xref="taxon:32630"
                     /clone="GH00329L1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
                     /lab_host="DH5alpha T1 resistant"
                     /note="Vector: pDNR-Dual"
```

```
     CDS             1. .>618
                     /note="Mutations: 617:Stop->Leu"
                     /codon_start=1
                     /transl_table=11
                     /product="Homo sapiens regulator of G-protein signalling
                     4"
                     /protein_id="AAP36424.1"
                     /db_xref="GI:30584351"
                     /translation="MCKGLAGLPASCLRSAKDMKHRLGFLLQKSDSCEHNSSHNKKDK
                     VVICQRVSQEEVKKWAESLENLISHECGLAAFKAFLKSEYSEENIDFWISCEEYKKIK
                     SPSKLSPKAKKIYNEFISVQATKEVNLDSCTREETSRNMLEPTITCFDEAQKKIFNLM
                     EKDSYRRFLKSRFYLDLVNPSSCGAEKQKGAKSSADCASLVPQCAL"
BASE COUNT      189 a    133 c    151 g    145 t
ORIGIN

  Query Match             22.6%;  Score 113;  DB 12;  Length 618;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-22;
  Matches  113;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 436

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 501
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 489


RESULT 10
AR270528
LOCUS       AR270528                800 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 1091 from patent US 6500938.
ACCESSION   AR270528
VERSION     AR270528.1  GI:29701762
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   Au-Young,J. and Seilhamer,J.J.
  TITLE     Composition for the detection of signaling pathway gene expression
  JOURNAL   Patent: US 6500938-A 1091 31-DEC-2002;
FEATURES             Location/Qualifiers
     source          1. .800
                     /organism="unknown"
BASE COUNT      241 a    181 c    195 g    183 t
ORIGIN

  Query Match             22.6%;  Score 113;  DB 6;  Length 800;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-22;
  Matches  113;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        474 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 533

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 501
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        534 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 586


RESULT 11
HSU27768
LOCUS       HSU27768                800 bp    mRNA    linear   PRI 07-MAR-1996
DEFINITION  Human RGP4 mRNA, complete cds.
ACCESSION   U27768
VERSION     U27768.1  GI:1216372
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   Druey,K.M., Blumer,K.J., Kang,V.H. and Kehrl,J.H.
  TITLE     Inhibition of G-protein-mediated MAP kinase activation by a new
            mammalian gene family
  JOURNAL   Nature 379 (6567), 742-746 (1996)
  MEDLINE   96178495
   PUBMED   8602223
REFERENCE   2  (bases 1 to 800)
  AUTHORS   Druey,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-1995) Kirk Druey, Intramural Research/NIAID/LIR,
            Rm 11B13, National Institutes of Health, 10 Center Drive, MSC 1876,
            Bethesda, MD 20892-1876, USA
FEATURES             Location/Qualifiers
     source          1. .800
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="brain"
     CDS             98. .715
                     /codon_start=1
                     /product="RGP4"
                     /protein_id="AAC50395.1"
                     /db_xref="GI:1216373"
                     /translation="MCKGLAGLPASCLRSAKDMKHRLGFLLQKSDSCEHNSSHNKKDK
                     VVICQRVSQEEVKKWAESLENLISHECGLAAFKAFLKSEYSEENIDFWISCEEYKKIK
                     SPSKLSPKAKKIYNEFISVQATKEVNLDSCTREETSRNMLEPTITCFDEAQKKIFNLM
                     EKDSYRRFLKSRFYLDLVNPSSCGAEKQKGAKSSADCASLVPQCA"
BASE COUNT      241 a    181 c    195 g    183 t
ORIGIN

  Query Match             22.6%;  Score 113;  DB 9;  Length 800;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-22;
  Matches  113;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        474 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 533

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 501
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        534 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 586


RESULT 12
BC000737
LOCUS       BC000737                1869 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, regulator of G-protein signalling 4, clone MGC:2124
            IMAGE:3510260, mRNA, complete cds.
ACCESSION   BC000737
VERSION     BC000737.1  GI:12653888
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1869)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan
```

```
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 7 Row: d Column: 22
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 1216372.
FEATURES              Location/Qualifiers
     source           1. .1869
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="LocusID:5999"
                      /db_xref="taxon:9606"
                      /clone="MGC:2124 IMAGE:3510260"
                      /tissue_type="Brain, neuroblastoma"
                      /clone_lib="NIH_MGC_19"
                      /lab_host="DH10B-R"
                      /note="Vector: pOTB7"
     CDS              86. .703
                      /codon_start=1
                      /product="regulator of G-protein signalling 4"
                      /protein_id="AAH00737.1"
                      /db_xref="GI:12653889"
                      /translation="MCKGLAGLPASCLRSAKDMKHRLGFLLQKSDSCEHNSSHNKKDK
                      VVICQRVSQEEVKKWAESLENLISHECGLAAFKAFLKSEYSEENIDFWISCEEYKKIK
                      SPSKLSPKAKKIYNEFISVQATKEVNLDSCTREETSRNMLEPTITCFDEAQKKIFNLM
                      EKDSYRRFLKSRFYLDLVNPSSCGAEKQKGAKSSADCASLVPQCA"
BASE COUNT      531 a    392 c    418 g    528 t
ORIGIN

  Query Match             22.6%;  Score 113;  DB 9;  Length 1869;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-22;
  Matches  113;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        462 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 521

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 501
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        522 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 574


RESULT 13
BC051869
LOCUS       BC051869                2753 bp    mRNA    linear   PRI 05-MAY-2003
DEFINITION  Homo sapiens regulator of G-protein signalling 4, mRNA (cDNA clone
            MGC:60244 IMAGE:5786695), complete cds.
ACCESSION   BC051869
VERSION     BC051869.1  GI:30354011
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2753)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2753)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 110 Row: g Column: 22
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 11184227.
FEATURES              Location/Qualifiers
     source           1. .2753
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="MGC:60244 IMAGE:5786695"
                      /tissue_type="Uterus, leiomyosarcoma"
                      /clone_lib="NIH_MGC_71"
                      /lab_host="DH10B"
                      /note="Vector: pCMV-SPORT6"
     gene             1. .2753
                      /gene="RGS4"
                      /db_xref="LocusID:5999"
                      /db_xref="MIM:602516"
     CDS              98. .715
                      /codon_start=1
                      /product="RGS4 protein"
                      /protein_id="AAH51869.1"
                      /db_xref="GI:30354012"
                      /db_xref="LocusID:5999"
                      /translation="MCKGLAGLPASCLRSAKDMKHRLGFLLQKSDSCEHNSSHNKKDK
                      VVICQRVSQEEVKKWAESLENLISHECGLAAFKAFLKSEYSEENIDFWISCEEYKKIK
                      SPSKLSPKAKKIYNEFISVQATKEVNLDSCTREETSRNMLEPTITCFDEAQKKIFNLM
                      EKDSYRRFLKSRFYLDLVNPSSCGAEKQKGAKSSADCASLVPQCA"
BASE COUNT      784 a    549 c    570 g    850 t
ORIGIN

  Query Match             22.6%;  Score 113;  DB 9;  Length 2753;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-22;
  Matches  113;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        474 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 533

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 501
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        534 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 586


RESULT 14
AX451335
LOCUS       AX451335                2934 bp    DNA     linear   PAT 03-JUL-2002
```

```
DEFINITION  Sequence 1 from Patent WO0216653.
ACCESSION   AX451335
VERSION     AX451335.1  GI:21698387
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Levitt,P.R., Mirnics,K., Kodavali,V.C. and Nimgaonkar,V.L.
  TITLE     Methods and systems for facilitating the diagnosis and treatment of
            schizophrenia
  JOURNAL   Patent: WO 0216653-A 1 28-FEB-2002;
            University of Pittsburgh (US)
FEATURES             Location/Qualifiers
     source          1. .2934
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      818 a    594 c    599 g    923 t
ORIGIN

  Query Match             22.6%;  Score 113;  DB 6;  Length 2934;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-22;
  Matches  113;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        453 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 512

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 501
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        513 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 565


RESULT 15
AK093959
LOCUS       AK093959               2599 bp    mRNA    linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ36640 fis, clone TRACH2019151, moderately
            similar to REGULATOR OF G-PROTEIN SIGNALING 4.
ACCESSION   AK093959
VERSION     AK093959.1  GI:21752924
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Suzuki,O., Sasaki,N., Aotsuka,S., Shoji,T., Ichihara,T.,
            Shiohata,N., Matsumoto,K., Hirano,M., Sano,S., Nomura,R.,
            Yoshikawa,Y., Matsumura,Y., Moriya,S., Chiba,E., Momiyama,H.,
            Onogawa,S., Kaeriyama,S., Satoh,N., Matsunawa,H., Takahashi,E.,
            Kataoka,R., Kuga,N., Kuroda,A., Satoh,I., Kamata,K., Takami,S.,
            Terashima,Y., Watanabe,M., Sugiyama,T., Irie,R., Otsuki,T.,
            Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2599)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1. .2599
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="TRACH2019151"
                     /tissue_type="trachea"
                     /clone_lib="TRACH2"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      719 a    520 c    541 g    819 t
ORIGIN

  Query Match             22.3%;  Score 111.8;  DB 9;  Length 2599;
  Best Local Similarity   94.3%;  Pred. No. 3.3e-22;
  Matches  116;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        379 TTTGCCCCTCAGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTA 438
              || | |    | ||||||||||||||||||||||||||||||||||||||||||||||
Db        324 TTAGTCATGAATGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTA 383

Qy        439 GAGCCTACAATAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAG 498
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        384 GAGCCTACAATAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAG 443

Qy        499 GAT 501
              |||
Db        444 GAT 446


Search completed: November  7, 2003, 11:11:24
Job time : 1967.4 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 7, 2003, 01:54:16 ; Search time 1962.4 Seconds
(without alignments)
10444.222 Million cell updates/sec

Title: US-09-939-209A-3_COPY_19800_20300
Perfect score: 501
Sequence: 1 ccacatgattatctcaatag.....atgagtgaactcccattcac 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
	1	501	100.0	20300	6	AX451337	AX451337 Sequence
	2	501	100.0	165329	9	AL583850	AL583850 Human DNA
c	3	496.2	99.0	191699	2	AC031977	AC031977 Homo sapi
c	4	481.8	96.2	116800	9	AL391379	AL391379 Human DNA
	5	480.2	95.8	165295	2	AC068486	AC068486 Homo sapi
	6	480.2	95.8	194867	2	AC023152	AC023152 Homo sapi
c	7	478.6	95.5	149428	2	AC010264	AC010264 Homo sapi
c	8	478.6	95.5	163314	2	AL590557	AL590557 Homo sapi
	9	478.6	95.5	173158	9	AC090096	AC090096 Homo sapi
	10	478.6	95.5	177097	2	AP001569	AP001569 Homo sapi
	11	478.6	95.5	179726	9	AC007052	AC007052 Homo sapi
	12	478.6	95.5	182411	2	AC090408	AC090408 Homo sapi
c	13	478.6	95.5	186780	9	AC005740	AC005740 Homo sapi
c	14	478.6	95.5	196869	2	AC087535	AC087535 Homo sapi
c	15	478.6	95.5	199275	9	AC011401	AC011401 Homo sapi
	16	478.6	95.5	200774	2	AP001592	AP001592 Homo sapi
c	17	477	95.2	21399	9	AC131012	AC131012 Homo sapi
c	18	477	95.2	41389	9	AP000542	AP000542 Homo sapi
	19	477	95.2	48103	9	AL356282	AL356282 Human DNA
c	20	477	95.2	60451	9	AL392112	AL392112 Human DNA
c	21	477	95.2	77127	9	AC104065	AC104065 Homo sapi
	22	477	95.2	78735	9	AC006545	AC006545 Homo sapi
	23	477	95.2	84579	2	AL356380	AL356380 Homo sapi
c	24	477	95.2	94882	2	AC002317	AC002317 Homo sapi
	25	477	95.2	97352	9	AL441885	AL441885 Human DNA
c	26	477	95.2	97771	9	BX088563	BX088563 Human DNA
	27	477	95.2	109140	2	AP001969	AP001969 Homo sapi
	28	477	95.2	112084	9	AC104648	AC104648 Homo sapi
c	29	477	95.2	116879	9	AC106854	AC106854 Homo sapi
	30	477	95.2	132337	9	AC104798	AC104798 Homo sapi
	31	477	95.2	136198	9	AC092092	AC092092 Homo sapi
	32	477	95.2	142080	9	AC079748	AC079748 Homo sapi
c	33	477	95.2	144136	2	AC012243	AC012243 Homo sapi
	34	477	95.2	153624	9	CNS01DS4	AL121654 BAC seque
	35	477	95.2	158091	2	AC025255	AC025255 Homo sapi
c	36	477	95.2	159698	9	AC027269	AC027269 Homo sapi
c	37	477	95.2	162746	9	AC006050	AC006050 Homo sapi
c	38	477	95.2	165047	9	AC079835	AC079835 Homo sapi
c	39	477	95.2	172579	9	AC036125	AC036125 Homo sapi
c	40	477	95.2	172600	9	AL359644	AL359644 Human DNA
c	41	477	95.2	173728	9	AL365496	AL365496 Human DNA
	42	477	95.2	176597	2	AC080190	AC080190 Homo sapi
	43	477	95.2	177176	2	AC040995	AC040995 Homo sapi
c	44	477	95.2	180964	9	AC007251	AC007251 Homo sapi
	45	477	95.2	189316	9	CNS01DUK	AL133247 BAC seque

ALIGNMENTS

RESULT 1
AX451337
LOCUS AX451337 20300 bp DNA linear PAT 03-JUL-2002
DEFINITION Sequence 3 from Patent WO0216653.
ACCESSION AX451337
VERSION AX451337.1 GI:21698388
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Levitt,P.R., Mirnics,K., Kodavali,V.C. and Nimgaonkar,V.L.
TITLE Methods and systems for facilitating the diagnosis and treatment of schizophrenia
JOURNAL Patent: WO 0216653-A 3 28-FEB-2002;

```
            University of Pittsburgh (US)
FEATURES             Location/Qualifiers
     source          1. .20300
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="A genomic sequence containing RGS4 nucleic acid
                     sequence and sequ ences upstream and downstream to the
                     RGS4 nucleic acid sequence"
BASE COUNT     6157 a   4102 c   3775 g   6266 t
ORIGIN

  Query Match           100.0%;  Score 501;  DB 6;  Length 20300;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-114;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19800 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 19859

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19860 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 19919

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19920 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 19979

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19980 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 20039

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20040 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 20099

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20100 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 20159

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20160 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 20219

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20220 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 20279

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db      20280 CATGAGTGAACTCCCATTCAC 20300


RESULT 2
AL583850
LOCUS       AL583850               165329 bp    DNA     linear   PRI 15-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-430G6 on chromosome 1, complete
            sequence.
ACCESSION   AL583850
VERSION     AL583850.5  GI:16973044
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165329)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 16, 2001 this sequence version replaced gi:15020514.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-430G6 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-430G6 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP11-430G6 is at 165329 in this
            sequence. The true right end of clone RP11-331H2 is at 2000 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .165329
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-430G6"
                     /clone_lib="RPCI-11.2"
BASE COUNT    51144 a  30897 c  31439 g  51849 t
ORIGIN

  Query Match           100.0%;  Score 501;  DB 9;  Length 165329;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-114;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91481 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 91540

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91541 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 91600

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91601 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 91660

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91661 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 91720

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91721 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 91780

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91781 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 91840

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91841 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 91900
```

```
Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91901 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 91960

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db      91961 CATGAGTGAACTCCCATTCAC 91981


RESULT 3
AC031977/c
LOCUS       AC031977              191699 bp    DNA     linear   HTG 12-APR-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-288O18, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC031977
VERSION     AC031977.7  GI:13194952
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
            Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
            Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
            Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Mar 4, 2001 this sequence version replaced gi:9665085.
            -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 880
                 Center clone name: RP11-288O18
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513; 100% of reads
                 Sequencing Vector: plasmid; plasmid_accession; 0% of reads
                 Chemistry: Dye-primer; 1% of reads
                 Chemistry: Dye-terminator Big Dye; 99% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 190680 bases at least Q40
                 Consensus quality: 191287 bases at least Q30
                 Consensus quality: 191336 bases at least Q20
                 Insert size: 195548; agarose-fp
                 Insert size: 191499; sum-of-contigs
                 Quality coverage:  7.9x in Q20 bases; agarose-fp
                 Quality coverage:  8.1x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    12646: contig of 12646 bp in length
            *    12647    12746: gap of unknown length
            *    12747    94961: contig of 82215 bp in length
            *    94962    95061: gap of unknown length
            *    95062   191699: contig of 96638 bp in length
FEATURES             Location/Qualifiers
     source          1. .191699
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-288O18"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .12646
                     /note="assembly_name:Contig5"
     misc_feature    12747. .94961
                     /note="assembly_name:Contig6
                     clone_end:T7"
     misc_feature    95062. .191699
                     /note="assembly_name:Contig7
                     clone_end:SP6"
BASE COUNT    57654 a  36385 c  36166 g  61293 t    201 others
ORIGIN

  Query Match             99.0%;  Score 496.2;  DB 2;  Length 191699;
  Best Local Similarity   99.4%;  Pred. No. 8.8e-113;
  Matches  498;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              |||||||||||||||||||||||||||||||| |||||||||||| ||||||||||||||
Db     184686 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACTCTTCAT 184627

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     184626 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 184567

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     184566 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 184507

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     184506 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 184447

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     184446 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 184387

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     184386 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 184327

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     184326 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 184267

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     184266 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 184207

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db     184206 CATGAGTGAACTCCCATTCAC 184186


RESULT 4
AL391379/c
LOCUS       AL391379              116800 bp    DNA     linear   PRI 05-APR-2001
DEFINITION  Human DNA sequence from clone RP13-171J5 on chromosome X, complete
            sequence.
ACCESSION   AL391379
VERSION     AL391379.12  GI:13560021
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 116800)
  AUTHORS   Chapman,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 6, 2001 this sequence version replaced gi:13446478.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP13-171J5 is from the library RPCI-13.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP13-171J5 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP13-171J5 is at 116800 in this
            sequence. The true right end of clone RP11-348F1 is at 100 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .116800
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP13-171J5"
                     /clone_lib="RPCI-13.1"
     repeat_region   135. .1785
                     /note="L1PA4 repeat: matches 4494. .6146 of consensus"
     repeat_region   1924. .1979
                     /note="L2 repeat: matches 2693. .2748 of consensus"
     repeat_region   2397. .2867
                     /note="L1ME1 repeat: matches 5693. .6164 of consensus"
     repeat_region   2888. .2954
                     /note="THE1B repeat: matches 328. .394 of consensus"
     repeat_region   2965. .3777
                     /note="L1PA13 repeat: matches -651. .189 of consensus"
     repeat_region   3803. .4094
                     /note="L1PA13 repeat: matches 288. .606 of consensus"
     repeat_region   4109. .4774
                     /note="L1PA13 repeat: matches 758. .1353 of consensus"
     repeat_region   4775. .5070
                     /note="AluSx repeat: matches 1. .306 of consensus"
     repeat_region   5071. .5352
                     /note="L1PA13 repeat: matches 1353. .1632 of consensus"
     repeat_region   5348. .8788
                     /note="L1P repeat: matches 1794. .5302 of consensus"
     repeat_region   8789. .8916
                     /note="MSTB repeat: matches 1. .112 of consensus"
     repeat_region   8919. .9466
                     /note="L1ME repeat: matches 5090. .5675 of consensus"
     repeat_region   9528. .9653
                     /note="MLT1H repeat: matches 319. .451 of consensus"
     repeat_region   9663. .9679
                     /note="L1MA3 repeat: matches 6289. .6304 of consensus"
     repeat_region   9680. .9995
                     /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   9996. .10726
                     /note="L1MA3 repeat: matches 5527. .6289 of consensus"
     repeat_region   10728. .11133
                     /note="L1ME repeat: matches 4887. .5298 of consensus"
     repeat_region   11130. .11545
                     /note="L1M1 repeat: matches 2929. .3347 of consensus"
     repeat_region   11546. .11841
                     /note="AluSq repeat: matches 1. .294 of consensus"
     repeat_region   11842. .13729
                     /note="L1M1 repeat: matches 1029. .2929 of consensus"
     repeat_region   13730. .14034
                     /note="AluSc repeat: matches 1. .305 of consensus"
     repeat_region   14035. .16479
                     /note="L1M1 repeat: matches -1380. .1029 of consensus"
     repeat_region   16965. .17477
                     /note="L1MA2 repeat: matches 5754. .6308 of consensus"
     repeat_region   17481. .18227
                     /note="L1PA5 repeat: matches 5401. .6143 of consensus"
     repeat_region   19334. .19440
                     /note="MIR repeat: matches 53. .166 of consensus"
     misc_feature    19997. .20454
                     /note="match: STS: Em:HSC11C12"
     repeat_region   20264. .20588
                     /note="L2 repeat: matches 2179. .2500 of consensus"
     repeat_region   20748. .20880
                     /note="L2 repeat: matches 2617. .2750 of consensus"
     repeat_region   21429. .21593
                     /note="MIR repeat: matches 46. .214 of consensus"
     misc_feature    complement(21548. .21900)
                     /note="match: STS: Em:G51507
                     match: GSS: Em:AQ198449"
     misc_feature    complement(21674. .21943)
                     /note="match: GSS: Em:AQ267386"
     repeat_region   22601. .22780
                     /note="MIR repeat: matches 45. .230 of consensus"
     repeat_region   23742. .25361
                     /note="L1PA6 repeat: matches 4524. .6143 of consensus"
     repeat_region   25363. .25437
                     /note="L1PA3 repeat: matches 6071. .6146 of consensus"
     repeat_region   25652. .25803
                     /note="L2 repeat: matches 2567. .2746 of consensus"
     repeat_region   26117. .26189
                     /note="MIR repeat: matches 188. .262 of consensus"
     repeat_region   27555. .29894
                     /note="L1MC3 repeat: matches 5355. .7736 of consensus"
     repeat_region   29901. .30357
                     /note="L1M4 repeat: matches 2942. .3407 of consensus"
     repeat_region   30404. .30525
                     /note="61 copies 2 mer tt 59% conserved"
     repeat_region   30591. .30881
                     /note="AluSx repeat: matches 1. .291 of consensus"
     repeat_region   31006. .31259
                     /note="L1MEc repeat: matches 1864. .2123 of consensus"
     repeat_region   31279. .31412
                     /note="L1MA5 repeat: matches 6155. .6290 of consensus"
     repeat_region   31389. .32051
                     /note="L1M4c repeat: matches 1271. .1908 of consensus"
     repeat_region   32093. .32293
                     /note="L1M4c repeat: matches 979. .1176 of consensus"
     repeat_region   33266. .33600
                     /note="L1MD repeat: matches 57. .398 of consensus"
     repeat_region   33678. .34684
                     /note="SVA repeat: matches 16. .955 of consensus"
     misc_feature    34139. .35218
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   34685. .35393
                     /note="SVA repeat: matches 724. .1386 of consensus"
     misc_feature    complement(35410. .35561)
```

```
                     /note="match: GSS: Em:AQ451532"
     misc_feature    35436. .35919
                     /note="match: GSS: Em:AQ004979"
     repeat_region   35580. .35797
                     /note="L1ME3A repeat: matches 5943. .6159 of consensus"
     repeat_region   36375. .36760
                     /note="L1MB7 repeat: matches 5787. .6170 of consensus"
     repeat_region   38032. .38055
                     /note="12 copies 2 mer ag 95% conserved"
     repeat_region   38425. .38829
                     /note="MLT2B repeat: matches 1. .409 of consensus"
     repeat_region   38855. .39078
                     /note="112 copies 2 mer tt 62% conserved"
     repeat_region   39086. .39120
                     /note="MLT2E repeat: matches 350. .386 of consensus"
     repeat_region   40118. .40157
                     /note="20 copies 2 mer tg 95% conserved"
     repeat_region   40228. .40339
                     /note="56 copies 2 mer at 72% conserved"
     repeat_region   40362. .40710
                     /note="MLT2E repeat: matches 3. .349 of consensus"
     misc_feature    41656. .42060
                     /note="match: GSS: Em:AQ150799"
     repeat_region   42170. .42372
                     /note="LTR41 repeat: matches 4. .188 of consensus"
     repeat_region   42797. .44565
                     /note="L1MB7 repeat: matches 4002. .5792 of consensus"
     repeat_region   44560. .45102
                     /note="L1M4 repeat: matches 2254. .2799 of consensus"
     repeat_region   45155. .45249
                     /note="L1M4 repeat: matches 4617. .4711 of consensus"
     repeat_region   45421. .45834
                     /note="L1M4 repeat: matches 1681. .2110 of consensus"
     repeat_region   45858. .46504
                     /note="L1M4c repeat: matches 1311. .1972 of consensus"
     repeat_region   46491. .52551
                     /note="L1PA3 repeat: matches 74. .6146 of consensus"
     repeat_region   52551. .52907
                     /note="L1M4 repeat: matches 998. .1354 of consensus"
     repeat_region   52901. .53112
                     /note="L1M4 repeat: matches -9. .205 of consensus"
     repeat_region   53251. .55347
                     /note="L1PA3 repeat: matches 4063. .6146 of consensus"
     misc_feature    56112. .56613

  Query Match             96.2%;  Score 481.8;  DB 9;  Length 116800;
  Best Local Similarity   97.6%;  Pred. No. 3.4e-109;
  Matches  489;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||| |||||||||||| ||||||||||||
Db      48744 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACTCTTCAT 48685

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              ||||||||||||||| |||||||||||||||||||||||||||||||||  ||||
Db      48684 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 48625

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      48624 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 48565

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48564 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 48505

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db      48504 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 48445

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      48444 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGT 48385

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48384 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 48325

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||||| |||||||| ||||||||||||||||||||||||||||
Db      48324 CAATGTACAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 48265

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db      48264 CATGAGTGAACTCCCATTCAC 48244


RESULT 5
AC068486
LOCUS       AC068486              165295 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome X clone RP11-316O5, WORKING DRAFT SEQUENCE,
            13 unordered pieces.
ACCESSION   AC068486
VERSION     AC068486.1  GI:7677976
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165295)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165295)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT
            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0316O05
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 160404 bases at least Q40
            Consensus quality: 162252 bases at least Q30
            Consensus quality: 163099 bases at least Q20
            Insert size: 175000; agarose-fp
            Insert size: 164095; sum-of-contigs
            Quality coverage: 4.79 in Q20 bases; agarose-fp
            Quality coverage: 5.12 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1758: contig of 1758 bp in length
            *     1759     1858: gap of unknown length
            *     1859     3835: contig of 1977 bp in length
            *     3836     3935: gap of unknown length
            *     3936     6268: contig of 2333 bp in length
            *     6269     6368: gap of unknown length
            *     6369     9527: contig of 3159 bp in length
            *     9528     9627: gap of unknown length
            *     9628    13500: contig of 3873 bp in length
```

```
            *    13501    13600: gap of unknown length
            *    13601    19166: contig of 5566 bp in length
            *    19167    19266: gap of unknown length
            *    19267    25083: contig of 5817 bp in length
            *    25084    25183: gap of unknown length
            *    25184    34033: contig of 8850 bp in length
            *    34034    34133: gap of unknown length
            *    34134    46088: contig of 11955 bp in length
            *    46089    46188: gap of unknown length
            *    46189    65638: contig of 19450 bp in length
            *    65639    65738: gap of unknown length
            *    65739    93577: contig of 27839 bp in length
            *    93578    93677: gap of unknown length
            *    93678   120965: contig of 27288 bp in length
            *   120966   121065: gap of unknown length
            *   121066   165295: contig of 44230 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165295
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-316O5"
     misc_feature    1. .1758
                     /note="assembly_name:Contig3"
     misc_feature    1859. .3835
                     /note="assembly_name:Contig4
                     clone_end:T7
                     vector_side:right"
     misc_feature    3936. .6268
                     /note="assembly_name:Contig5"
     misc_feature    6369. .9527
                     /note="assembly_name:Contig6"
     misc_feature    9628. .13500
                     /note="assembly_name:Contig7"
     misc_feature    13601. .19166
                     /note="assembly_name:Contig8"
     misc_feature    19267. .25083
                     /note="assembly_name:Contig9
                     clone_end:SP6
                     vector_side:left"
     misc_feature    25184. .34033
                     /note="assembly_name:Contig10"
     misc_feature    34134. .46088
                     /note="assembly_name:Contig11"
     misc_feature    46189. .65638
                     /note="assembly_name:Contig12"
     misc_feature    65739. .93577
                     /note="assembly_name:Contig13"
     misc_feature    93678. .120965
                     /note="assembly_name:Contig14"
     misc_feature    121066. .165295
                     /note="assembly_name:Contig15"
BASE COUNT    50879 a  32297 c  32077 g  48840 t   1202 others
ORIGIN

  Query Match             95.8%;  Score 480.2;  DB 2;  Length 165295;
  Best Local Similarity   97.4%;  Pred. No. 8.1e-109;
  Matches  488;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              |||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db      83733 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACGCTTCAT 83792

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              ||||||||||||||| ||||||||||||||||||||||||||||||||||  ||||
Db      83793 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 83852

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      83853 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 83912

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83913 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 83972

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db      83973 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 84032

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      84033 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGT 84092

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84093 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 84152

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||
Db      84153 CAATGTACAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 84212

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db      84213 CATGAGTGAACTCCCATTCAC 84233


RESULT 6
AC023152
LOCUS       AC023152               194867 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome X clone RP11-737B10, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC023152
VERSION     AC023152.3  GI:7191122
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194867)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 194867)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 7, 2000 this sequence version replaced gi:7140268.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0737B10
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 192787 bases at least Q40
            Consensus quality: 193180 bases at least Q30
            Consensus quality: 193457 bases at least Q20
            Insert size: 190000; agarose-fp
            Insert size: 194667; sum-of-contigs
            Quality coverage: 7.06 in Q20 bases; agarose-fp
            Quality coverage: 6.92 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
```

```
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     8464: contig of 8464 bp in length
            *     8465     8564: gap of unknown length
            *     8565    78379: contig of 69815 bp in length
            *    78380    78479: gap of unknown length
            *    78480   194867: contig of 116388 bp in length.
FEATURES             Location/Qualifiers
     source          1. .194867
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-737B10"
     misc_feature    1. .8464
                     /note="assembly_name:Contig4
                     clone_end:T7
                     vector_side:right"
     misc_feature    8565. .78379
                     /note="assembly_name:Contig5
                     clone_end:SP6
                     vector_side:right"
     misc_feature    78480. .194867
                     /note="assembly_name:Contig6"
BASE COUNT    60197 a  38604 c  39004 g  56862 t    200 others
ORIGIN

  Query Match             95.8%;  Score 480.2;  DB 2;  Length 194867;
  Best Local Similarity   97.4%;  Pred. No. 8e-109;
  Matches  488;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              |||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db     126543 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACGCTTCAT 126602

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              ||||||||||||||| ||||||||||||||||||||||||||||||||||  ||||
Db     126603 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 126662

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     126663 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 126722

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     126723 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 126782

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     126783 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 126842

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     126843 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGT 126902

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     126903 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 126962

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||
Db     126963 CAATGTACAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 127022

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db     127023 CATGAGTGAACTCCCATTCAC 127043


RESULT 7
AC010264/c
LOCUS       AC010264               149428 bp    DNA     linear   HTG 18-JUL-2000
DEFINITION  Homo sapiens chromosome 5 clone CTC-468K18, WORKING DRAFT SEQUENCE,
            23 ordered pieces.
ACCESSION   AC010264
VERSION     AC010264.5  GI:9256174
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 149428)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 149428)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 18, 2000 this sequence version replaced gi:7710802.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 365150
            Center clone name: CIT-HSPC_468K18
            -------
            Summary Statistics
            Consensus quality: 136337 bases at least Q40
            Consensus quality: 144306 bases at least Q30
            Consensus quality: 146065 bases at least Q20
            Estimated insert size: 150000; pulse field gel estimation
            Estimated insert size: 148378; sum-of-contigs estimation
            Quality coverage:  4.59 in Q20 bases; pulse field gel estimation
            Quality coverage:  4.64 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     8196: contig of 8196 bp in length
            *     8197     8296: gap of unknown length
            *     8297    12916: contig of 4620 bp in length
            *    12917    13016: gap of unknown length
            *    13017    14756: contig of 1740 bp in length
            *    14757    14856: gap of unknown length
            *    14857    19143: contig of 4287 bp in length
            *    19144    19243: gap of unknown length
            *    19244    20796: contig of 1553 bp in length
            *    20797    20896: gap of unknown length
            *    20897    28283: contig of 7387 bp in length
            *    28284    28383: gap of unknown length
            *    28384    43450: contig of 15067 bp in length
            *    43451    43550: gap of unknown length
            *    43551    55287: contig of 11737 bp in length
            *    55288    55387: gap of unknown length
            *    55388    73601: contig of 18214 bp in length
            *    73602    73701: gap of unknown length
            *    73702    78295: contig of 4594 bp in length
            *    78296    78395: gap of unknown length
            *    78396    81832: contig of 3437 bp in length
            *    81833    81932: gap of unknown length
            *    81933    83998: contig of 2066 bp in length
            *    83999    84098: gap of unknown length
            *    84099    86531: contig of 2433 bp in length
            *    86532    86631: gap of unknown length
            *    86632    88757: contig of 2126 bp in length
            *    88758    88857: gap of unknown length
            *    88858    95081: contig of 6224 bp in length
```

```
            *    95082    95181: gap of unknown length
            *    95182    98288: contig of 3107 bp in length
            *    98289    98388: gap of unknown length
            *    98389   107715: contig of 9327 bp in length
            *   107716   107815: gap of unknown length
            *   107816   112037: contig of 4222 bp in length
            *   112038   112137: gap of unknown length
            *   112138   114033: contig of 1896 bp in length
            *   114034   114133: gap of unknown length
            *   114134   119274: contig of 5141 bp in length
            *   119275   119374: gap of unknown length
            *   119375   126879: contig of 7505 bp in length
            *   126880   126979: gap of unknown length
            *   126980   129104: contig of 2125 bp in length
            *   129105   129204: gap of unknown length
            *   129205   149428: contig of 20224 bp in length.
FEATURES             Location/Qualifiers
     source          1. .149428
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTC-468K18"
                     /clone_lib="CalTech human BAC library C"
BASE COUNT    39562 a  32858 c  33987 g  40817 t   2204 others
ORIGIN

  Query Match             95.5%;  Score 478.6;  DB 2;  Length 149428;
  Best Local Similarity   97.2%;  Pred. No. 2e-108;
  Matches  487;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db      26772 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACCCTTCAT 26713

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              |||||||||||||||| ||||||||||||||||||||||||||||||||||  ||||
Db      26712 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 26653

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      26652 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 26593

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26592 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 26533

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db      26532 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 26473

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      26472 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGT 26413

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26412 CTCAGCGCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 26353

Qy        421 CAATGTACAAAAATCACAAGCACTCTTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||
Db      26352 CAATGTACAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 26293

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db      26292 CATGAGTGAACTCCCATTCAC 26272


RESULT 8
AL590557/c
LOCUS       AL590557               163314 bp    DNA     linear   HTG 20-JUL-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-24C8, *** SEQUENCING IN
            PROGRESS ***, 8 unordered pieces.
ACCESSION   AL590557
VERSION     AL590557.8  GI:13992136
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Mclay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 8, 2001 this sequence version replaced gi:13990622.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA24C8
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 161385 bases at least Q40
            Consensus quality: 161940 bases at least Q30
            Consensus quality: 162264 bases at least Q20
            Insert size: 162614; sum-of-contigs
            Insert size: 164357; 4.9% error; agarose-fp
            Quality coverage: 6.63x in Q20 bases; sum-of-contigs Quality
            coverage: 6.67x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    16091: contig of 16091 bp in length
            *    16092    16191: gap of 100 bp
            *    16192    51184: contig of 34993 bp in length
            *    51185    51284: gap of 100 bp
            *    51285    67941: contig of 16657 bp in length
            *    67942    68041: gap of 100 bp
            *    68042   103921: contig of 35880 bp in length
            *   103922   104021: gap of 100 bp
            *   104022   107020: contig of 2999 bp in length
            *   107021   107120: gap of 100 bp
            *   107121   123748: contig of 16628 bp in length
            *   123749   123848: gap of 100 bp
            *   123849   137091: contig of 13243 bp in length
            *   137092   137191: gap of 100 bp
            *   137192   163314: contig of 26123 bp in length.
FEATURES             Location/Qualifiers
     source          1. .163314
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-24C8"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .16091
                     /note="assembly_fragment:00603
                     fragment_chain:1"
     misc_feature    16192. .51184
                     /note="assembly_fragment:01328
                     fragment_chain:1"
     misc_feature    51285. .67941
                     /note="assembly_fragment:01579
                     fragment_chain:1"
```

```
     misc_feature    68042. .103921
                     /note="assembly_fragment:00704
                     fragment_chain:1"
     misc_feature    104022. .107020
                     /note="assembly_fragment:02225
                     fragment_chain:1"
     misc_feature    107121. .123748
                     /note="assembly_fragment:02172
                     fragment_chain:1"
     misc_feature    123849. .137091
                     /note="assembly_fragment:01998
                     fragment_chain:1"
     misc_feature    137192. .163314
                     /note="assembly_fragment:03121"
BASE COUNT    51215 a  31325 c  32024 g  48050 t    700 others
ORIGIN

  Query Match             95.5%;  Score 478.6;  DB 2;  Length 163314;
  Best Local Similarity   97.2%;  Pred. No. 2e-108;
  Matches  487;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              |||||||||||||||||||||||||||||||| |||||||||||| ||||||||||||
Db      53430 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACTCTTCAT 53371

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              |||||||||||||||| |||||||||||||||||||||||||||||||||||  ||||
Db      53370 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 53311

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||| |||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      53310 CTATTACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 53251

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||| |||||||
Db      53250 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATACTGTTGGA 53191

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db      53190 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 53131

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      53130 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 53071

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      53070 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 53011

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||||| ||||||||| ||| ||||||||||||||||||||||
Db      53010 CAATGTACAAAAATCACAAGCATTCTTATACACCAACAACAGACAAACAGAGAGCCAAAT 52951

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db      52950 CATGAGTGAACTCCCATTCAC 52930


RESULT 9
AC090096
LOCUS       AC090096               173158 bp    DNA     linear   PRI 28-FEB-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-159C14, complete sequence.
ACCESSION   AC090096
VERSION     AC090096.5  GI:18997379
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173158)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-159C14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173158)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 173158)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 173158)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
```

```
          Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL Submitted (28-FEB-2002) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Feb 28, 2002 this sequence version replaced gi:18874974.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          -------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
              Center project name: L12761
              Center clone name: 159_C_14
          --------------.
FEATURES             Location/Qualifiers
     source          1. .173158
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-159C14"
                     /clone_lib="RPCI-11 Human Male BAC"
     unsure          233. .236
                     /note="<30 qual SNGL region"
     repeat_region   351. .896
                     /rpt_family="LTR48B"
     repeat_region   complement(913. .1139)
                     /rpt_family="MER102a"
     repeat_region   1140. .1506
                     /rpt_family="MLT1A1"
     repeat_region   complement(1507. .1550)
                     /rpt_family="MER102a"
     unsure          1661. .1665
                     /note="<30 qual SNGL region"
     repeat_region   1762. .1851
                     /rpt_family="LTR40b"
     repeat_region   1852. .2178
                     /rpt_family="MER1B"
     unsure          1921. .1925
                     /note="<30 qual SNGL region"
     repeat_region   2179. .2514
                     /rpt_family="LTR40b"
     unsure          2196. .2201
                     /note="<30 qual SNGL region"
     repeat_region   2777. .3023
                     /rpt_family="MIR"
     repeat_region   4778. .5086
                     /rpt_family="MER89"
     repeat_region   6100. .6189
                     /rpt_family="L3"
     repeat_region   8312. .8340
                     /rpt_family="MER89"
     repeat_region   complement(8341. .8641)
                     /rpt_family="AluY"
     repeat_region   8642. .8688
                     /rpt_family="MER89"
     repeat_region   8675. .9046
                     /rpt_family="MER89"
     repeat_region   9047. .9334
                     /rpt_family="L3"
     repeat_region   10587. .10793
                     /rpt_family="MER34"
     repeat_region   10836. .11030
                     /rpt_family="MER34"
     repeat_region   11115. .11183
                     /rpt_family="CT-rich"
     repeat_region   complement(11973. .12973)
                     /rpt_family="THE1C-int"
     repeat_region   complement(12975. .13339)
                     /rpt_family="THE1C"
     repeat_region   13457. .13487
                     /rpt_family="AT_rich"
     repeat_region   13530. .13645
                     /rpt_family="AluJb"
     repeat_region   13646. .13668
                     /rpt_family="(TTTA)n"
     repeat_region   complement(13672. .13943)
                     /rpt_family="AluSx"
     repeat_region   13944. .14149
                     /rpt_family="AluJb"
     repeat_region   14168. .14227
                     /rpt_family="AT_rich"
     repeat_region   14459. .14769
                     /rpt_family="MER33"
     repeat_region   14822. .14854
                     /rpt_family="AT_rich"
     repeat_region   15112. .15183
                     /rpt_family="(CA)n"
     repeat_region   15228. .15321
                     /rpt_family="L2"
     repeat_region   complement(15635. .15750)
                     /rpt_family="L2"
     repeat_region   17971. .17998
                     /rpt_family="AT_rich"
     repeat_region   19616. .19637
                     /rpt_family="AT_rich"
     repeat_region   19671. .19981
                     /rpt_family="L2"
     repeat_region   19982. .20204
                     /rpt_family="AluJb"
     repeat_region   20208. .20308
                     /rpt_family="L2"
     repeat_region   complement(20315. .20471)
                     /rpt_family="MIR"
     repeat_region   complement(20649. .20947)
                     /rpt_family="AluSg"
     repeat_region   21020. .21377
                     /rpt_family="MLT1A1"
     repeat_region   21428. .21626
                     /rpt_family="MER20"
```

```
Query Match             95.5%;  Score 478.6;  DB 9;  Length 173158;
Best Local Similarity   97.2%;  Pred. No. 2e-108;
Matches  487;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              |||||||||||||||||||||||||||||||| |||||||||||| | |||| ||||||
Db      29384 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAGCAACCCTTCAT 29443

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              ||||||||||||||| |||||||||||||||||||||||||||||||| ||| |||||
Db      29444 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAACAAGAACTAT 29503

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db      29504 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 29563

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      29564 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 29623

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||||||||||||||||||||||||||||||||||| ||||
Db      29624 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 29683

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      29684 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGT 29743

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      29744 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 29803
```

```
Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||
Db      29804 CAATGTACAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 29863

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db      29864 CATGAGTGAACTCCCATTCAC 29884


RESULT 10
AP001569
LOCUS       AP001569               177097 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-859C21 map 18q21, WORKING
            DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION   AP001569
VERSION     AP001569.2  GI:8117403
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177097)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 177,097 genomic DNA of 18q21
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 177097)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 30, 2000 this sequence version replaced gi:7380904.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft18
                Center clone name: RP11-859C21
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 160670 bases at least Q40
                Consensus quality: 168058 bases at least Q30
                Consensus quality: 171631 bases at least Q20
                Insert size: 173997; sum-of-contigs
                Quality coverage: 4.63x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            32 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                 1     23792 contig of  23792 bp in length
             23893     45465 contig of  21573 bp in length
             45566     58167 contig of  12602 bp in length
             58268     71173 contig of  12906 bp in length
             71274     81253 contig of   9980 bp in length
             81354     91080 contig of   9727 bp in length
             91181     97387 contig of   6207 bp in length
             97488    104902 contig of   7415 bp in length
            105003    111107 contig of   6105 bp in length
            111208    116851 contig of   5644 bp in length
            116952    123585 contig of   6634 bp in length
            123686    128985 contig of   5300 bp in length
            129086    134233 contig of   5148 bp in length
            134334    138588 contig of   4255 bp in length
            138689    142573 contig of   3885 bp in length
            142674    146286 contig of   3613 bp in length
            146387    150053 contig of   3667 bp in length
            150154    154046 contig of   3893 bp in length
            154147    156225 contig of   2079 bp in length
            156326    158578 contig of   2253 bp in length
            158679    161124 contig of   2446 bp in length
            161225    163559 contig of   2335 bp in length
            163660    165538 contig of   1879 bp in length
            165639    166949 contig of   1311 bp in length
            167050    168240 contig of   1191 bp in length
            168341    169579 contig of   1239 bp in length
            169680    170978 contig of   1299 bp in length
            171079    172080 contig of   1002 bp in length
            172181    173514 contig of   1334 bp in length
            173615    174614 contig of   1000 bp in length
            174715    175976 contig of   1262 bp in length
            176077    177097 contig of   1021 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1    23792: contig of 23792 bp in length
*    23793    23892: gap of 100 bp
*    23893    45465: contig of 21573 bp in length
*    45466    45565: gap of 100 bp
*    45566    58167: contig of 12602 bp in length
*    58168    58267: gap of 100 bp
*    58268    71173: contig of 12906 bp in length
*    71174    71273: gap of 100 bp
*    71274    81253: contig of 9980 bp in length
*    81254    81353: gap of 100 bp
*    81354    91080: contig of 9727 bp in length
*    91081    91180: gap of 100 bp
*    91181    97387: contig of 6207 bp in length
*    97388    97487: gap of 100 bp
*    97488   104902: contig of 7415 bp in length
*   104903   105002: gap of 100 bp
*   105003   111107: contig of 6105 bp in length
*   111108   111207: gap of 100 bp
*   111208   116851: contig of 5644 bp in length
*   116852   116951: gap of 100 bp
*   116952   123585: contig of 6634 bp in length
*   123586   123685: gap of 100 bp
*   123686   128985: contig of 5300 bp in length
*   128986   129085: gap of 100 bp
*   129086   134233: contig of 5148 bp in length
*   134234   134333: gap of 100 bp
*   134334   138588: contig of 4255 bp in length
*   138589   138688: gap of 100 bp
*   138689   142573: contig of 3885 bp in length
*   142574   142673: gap of 100 bp
*   142674   146286: contig of 3613 bp in length
*   146287   146386: gap of 100 bp
*   146387   150053: contig of 3667 bp in length
*   150054   150153: gap of 100 bp
*   150154   154046: contig of 3893 bp in length
*   154047   154146: gap of 100 bp
*   154147   156225: contig of 2079 bp in length
*   156226   156325: gap of 100 bp
*   156326   158578: contig of 2253 bp in length
*   158579   158678: gap of 100 bp
*   158679   161124: contig of 2446 bp in length
*   161125   161224: gap of 100 bp
```

```
          *   161225   163559: contig of 2335 bp in length
          *   163560   163659: gap of 100 bp
          *   163660   165538: contig of 1879 bp in length
          *   165539   165638: gap of 100 bp
          *   165639   166949: contig of 1311 bp in length
          *   166950   167049: gap of 100 bp
          *   167050   168240: contig of 1191 bp in length
          *   168241   168340: gap of 100 bp
          *   168341   169579: contig of 1239 bp in length
          *   169580   169679: gap of 100 bp
          *   169680   170978: contig of 1299 bp in length
          *   170979   171078: gap of 100 bp
          *   171079   172080: contig of 1002 bp in length
          *   172081   172180: gap of 100 bp
          *   172181   173514: contig of 1334 bp in length
          *   173515   173614: gap of 100 bp
          *   173615   174614: contig of 1000 bp in length
          *   174615   174714: gap of 100 bp
          *   174715   175976: contig of 1262 bp in length
          *   175977   176076: gap of 100 bp
          *   176077   177097: contig of 1021 bp in length.
FEATURES             Location/Qualifiers
     source          1. .177097
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18q21"
                     /clone="RP11-859C21"
     misc_feature    1. .23792
                     /note="assembly_fragment"
     misc_feature    23893. .45465
                     /note="assembly_fragment"
     misc_feature    45566. .58167
                     /note="assembly_fragment"
     misc_feature    58268. .71173
                     /note="assembly_fragment"
     misc_feature    71274. .81253
                     /note="assembly_fragment"
     misc_feature    81354. .91080
                     /note="assembly_fragment"
     misc_feature    91181. .97387
                     /note="assembly_fragment"
     misc_feature    97488. .104902
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    105003. .111107
                     /note="assembly_fragment"
     misc_feature    111208. .116851
                     /note="assembly_fragment"
     misc_feature    116952. .123585
                     /note="assembly_fragment"
     misc_feature    123686. .128985
                     /note="assembly_fragment"
     misc_feature    129086. .134233
                     /note="assembly_fragment"
     misc_feature    134334. .138588
                     /note="assembly_fragment"
     misc_feature    138689. .142573
                     /note="assembly_fragment"
     misc_feature    142674. .146286
                     /note="assembly_fragment clone_end:T7 vector_side:left"
     misc_feature    146387. .150053
                     /note="assembly_fragment"
     misc_feature    150154. .154046
                     /note="assembly_fragment"
     misc_feature    154147. .156225

  Query Match             95.5%;  Score 478.6;  DB 2;  Length 177097;
  Best Local Similarity   97.2%;  Pred. No. 2e-108;
  Matches  487;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||

Db     147438 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACACTTCAT 147497

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||  ||||

Db     147498 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 147557

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||

Db     147558 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 147617

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| |||||||||||||||||||||||||||||||| ||||||||||||||||||

Db     147618 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCCATTCAACATAGTGTTGGA 147677

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||

Db     147678 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 147737

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||

Db     147738 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGT 147797

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     147798 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 147857

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||

Db     147858 CAATGTACAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 147917

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db     147918 CATGAGTGAACTCCCATTCAC 147938


RESULT 11
AC007052
LOCUS       AC007052               179726 bp    DNA     linear   PRI 23-MAR-1999
DEFINITION  Homo sapiens chromosome 18, clone hRPK.411_H_24, complete sequence.
ACCESSION   AC007052
VERSION     AC007052.4  GI:4510438
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179726)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone hRPK.411_H_24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179726)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 179726)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
```

Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (23-MAR-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Mar 24, 1999 this sequence version replaced gi:4432872.
All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.

FEATURES Location/Qualifiers

source	1. .179726
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="18"
	/map="18"
	/clone="hRPK.411_H_24"
	/clone_lib="RPCI-11 human BAC library"
repeat_region	898. .1014
	/rpt_family="L1MD3"
repeat_region	complement(1035. .1333)
	/rpt_family="AluSx"
repeat_region	2987. .3030
	/rpt_family="CT-rich"
repeat_region	3087. .3291
	/rpt_family="L1MC4"
repeat_region	4298. .4330
	/rpt_family="(CA)n"
repeat_region	5344. .5365
	/rpt_family="AT_rich"
repeat_region	complement(5552. .5724)
	/rpt_family="AluJ"
repeat_region	5731. .11847
	/rpt_family="L1PA2"
repeat_region	11857. .11891
	/rpt_family="(TAAA)n"
repeat_region	complement(11893. .12002)
	/rpt_family="AluJ"
unsure	11925. .11948
	/note="Single-stranded coverage."
repeat_region	12026. .12321
	/rpt_family="AluSg"
repeat_region	complement(12605. .12846)
	/rpt_family="MIR"
repeat_region	13352. .13660
	/rpt_family="AluSc"
repeat_region	complement(13741. .14128)
	/rpt_family="MLT1B"
repeat_region	14425. .14456
	/rpt_family="AT_rich"
repeat_region	14457. .14978
	/rpt_family="MLT1F"
repeat_region	15496. .15834
	/rpt_family="THE1B"
repeat_region	16325. .16628
	/rpt_family="AluYb8"
repeat_region	16753. .17086
	/rpt_family="MER7A"
repeat_region	17543. .17601
	/rpt_family="MIR"
repeat_region	18075. .18442
	/rpt_family="MLT1B"
repeat_region	complement(18735. .18882)
	/rpt_family="MIR"
repeat_region	complement(18948. .19083)
	/rpt_family="L1MB8"
repeat_region	complement(19190. .19426)
	/rpt_family="AluJo"
repeat_region	19431. .19750
	/rpt_family="L1ME3"
repeat_region	20434. .20555
	/rpt_family="MLT1A1"
repeat_region	20556. .20863
	/rpt_family="AluSx"
repeat_region	20864. .21080
	/rpt_family="MLT1A1"
repeat_region	21528. .21565
	/rpt_family="AT_rich"
repeat_region	21575. .21668
	/rpt_family="L2"
repeat_region	complement(21849. .21991)
	/rpt_family="MER5B"
repeat_region	23147. .23168
	/rpt_family="(TTTG)n"
repeat_region	complement(23169. .23384)
	/rpt_family="AluJo"
repeat_region	23385. .23419
	/rpt_family="AT_rich"
repeat_region	complement(23799. .23846)
	/rpt_family="L2"
repeat_region	complement(23947. .24060)
	/rpt_family="L2"
repeat_region	24066. .24206
	/rpt_family="BC200"
repeat_region	complement(24347. .24514)
	/rpt_family="MER21B"
repeat_region	24782. .25203
	/rpt_family="L1MB2"
repeat_region	25385. .26005
	/rpt_family="L1MB2"
repeat_region	26035. .26336
	/rpt_family="AluSx"
repeat_region	26432. .26604
	/rpt_family="L1ME"
repeat_region	26610. .26645
	/rpt_family="AT_rich"
repeat_region	complement(27113. .27422)
	/rpt_family="AluSp"
repeat_region	27865. .28190
	/rpt_family="AluSg"
repeat_region	28250. .28272
	/rpt_family="AT_rich"
repeat_region	28363. .28496
	/rpt_family="HAL1"
repeat_region	complement(28497. .28796)
	/rpt_family="AluSg"
repeat_region	28797. .28869
	/rpt_family="HAL1"
repeat_region	28993. .29027
	/rpt_family="AT_rich"
repeat_region	29107. .29134
	/rpt_family="(TATATG)n"
repeat_region	29151. .29182
	/rpt_family="(TG)n"
repeat_region	complement(29821. .30164)
	/rpt_family="MER7A"
repeat_region	complement(30236. .30522)
	/rpt_family="AluJo"
repeat_region	30576. .30723
	/rpt_family="MIR"
repeat_region	30927. .31271
	/rpt_family="MLT1K"
repeat_region	32382. .32437
	/rpt_family="(CA)n"
repeat_region	32480. .32502
	/rpt_family="(TTTG)n"

```
     repeat_region   complement(32503. .32751)
                     /rpt_family="AluSx"
     repeat_region   34312. .34476
                     /rpt_family="MIR"
     repeat_region   complement(34648. .34953)
                     /rpt_family="AluSx"
     repeat_region   complement(35550. .35947)
                     /rpt_family="THE1B"
     repeat_region   36048. .36070
                     /rpt_family="(TTTTA)n"
     repeat_region   complement(36072. .36366)
                     /rpt_family="AluJo"
     repeat_region   36525. .36587
                     /rpt_family="AT_rich"
     repeat_region   36792. .36849
                     /rpt_family="AT_rich"
     repeat_region   complement(36903. .37092)
                     /rpt_family="L1MA7"

  Query Match             95.5%;  Score 478.6;  DB 9;  Length 179726;
  Best Local Similarity   97.2%;  Pred. No. 2e-108;
  Matches  487;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              |||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db       9594 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACACTTCAT 9653

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              |||||||||||||||| ||||||||||||||||||||||||||||||||||   ||||
Db       9654 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 9713

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       9714 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 9773

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| |||||||||||||||||||||||||||||||| |||||||||||||||||||
Db       9774 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCCATTCAACATAGTGTTGGA 9833

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       9834 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 9893

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db       9894 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGT 9953

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9954 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 10013

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||
Db      10014 CAATGTACAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 10073

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db      10074 CATGAGTGAACTCCCATTCAC 10094


RESULT 12
AC090408
LOCUS       AC090408               182411 bp    DNA     linear   HTG 09-MAY-2001
DEFINITION  Homo sapiens chromosome 18 clone RP11-859C21 map 18, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC090408
VERSION     AC090408.2  GI:13357356
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182411)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-859C21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 182411)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 16, 2001 this sequence version replaced gi:12958044.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12718
                Center clone name: 859_C_21
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 177348 bases at least Q40
                Consensus quality: 179869 bases at least Q30
                Consensus quality: 180657 bases at least Q20
                Insert size: 181111; sum-of-contigs
                Quality coverage: 6.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      803: contig of 803 bp in length
            *      804      903: gap of 100 bp
            *      904     1787: contig of 884 bp in length
            *     1788     1887: gap of 100 bp
            *     1888     3003: contig of 1116 bp in length
            *     3004     3103: gap of 100 bp
            *     3104     5498: contig of 2395 bp in length
            *     5499     5598: gap of 100 bp
            *     5599     7587: contig of 1989 bp in length
            *     7588     7687: gap of 100 bp
            *     7688     9739: contig of 2052 bp in length
            *     9740     9839: gap of 100 bp
            *     9840    13254: contig of 3415 bp in length
            *    13255    13354: gap of 100 bp
            *    13355    17739: contig of 4385 bp in length
```

```
            *    17740    17839: gap of 100 bp
            *    17840    25230: contig of 7391 bp in length
            *    25231    25330: gap of 100 bp
            *    25331    32077: contig of 6747 bp in length
            *    32078    32177: gap of 100 bp
            *    32178    48716: contig of 16539 bp in length
            *    48717    48816: gap of 100 bp
            *    48817    68998: contig of 20182 bp in length
            *    68999    69098: gap of 100 bp
            *    69099   103010: contig of 33912 bp in length
            *   103011   103110: gap of 100 bp
            *   103111   182411: contig of 79301 bp in length.
FEATURES             Location/Qualifiers
     source          1. .182411
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-859C21"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .803
                     /note="assembly_fragment"
     misc_feature    904. .1787
                     /note="assembly_fragment"
     misc_feature    1888. .3003
                     /note="assembly_fragment"
     misc_feature    3104. .5498
                     /note="assembly_fragment"
     misc_feature    5599. .7587
                     /note="assembly_fragment"
     misc_feature    7688. .9739
                     /note="assembly_fragment"
     misc_feature    9840. .13254
                     /note="assembly_fragment"
     misc_feature    13355. .17739
                     /note="assembly_fragment"
     misc_feature    17840. .25230
                     /note="assembly_fragment"
     misc_feature    25331. .32077
                     /note="assembly_fragment"
     misc_feature    32178. .48716
                     /note="assembly_fragment"
     misc_feature    48817. .68998
                     /note="assembly_fragment"
     misc_feature    69099. .103010
                     /note="assembly_fragment"
     misc_feature    103111. .182411
                     /note="assembly_fragment"
BASE COUNT    56444 a  35140 c  35375 g  54140 t   1312 others
ORIGIN

  Query Match             95.5%;  Score 478.6;  DB 2;  Length 182411;
  Best Local Similarity   97.2%;  Pred. No. 2e-108;
  Matches  487;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db     136011 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACACTTCAT 136070

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              |||||||||||||||| |||||||||||||||||||||||||||||||||||   ||||
Db     136071 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 136130

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     136131 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 136190

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| |||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     136191 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCCATTCAACATAGTGTTGGA 136250

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     136251 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 136310

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     136311 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGT 136370

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136371 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 136430

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||||||| ||||||||| ||||||||||||||||||||||||
Db     136431 CAATGTACAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 136490

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db     136491 CATGAGTGAACTCCCATTCAC 136511


RESULT 13
AC005740/c
LOCUS       AC005740               186780 bp    DNA     linear   PRI 01-OCT-1998
DEFINITION  Homo sapiens chromosome 5p, BAC clone 50g21 (LBNL H154), complete
            sequence.
ACCESSION   AC005740
VERSION     AC005740.1  GI:3687210
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186780)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186780)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 186780)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1. .186780
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5p"
                     /clone="50g21"
                     /note="LBNL H154"
     repeat_region   complement(1. .74)
                     /rpt_family="Alu"
     repeat_region   868. .1134
                     /rpt_family="Alu"
     repeat_region   complement(1208. .1514)
                     /rpt_family="Alu"
     repeat_region   complement(3486. .3634)
                     /rpt_family="L1"
     repeat_region   3638. .3938
                     /rpt_family="Alu"
     repeat_region   3921. .3943
```

```
                /note="(A)23"
                /rpt_type=tandem
                /rpt_unit=A
repeat_region   complement(5195. .5272)
                /rpt_family="MIR2"
repeat_region   complement(5719. .6005)
                /rpt_family="Alu"
repeat_region   complement(6469. .6637)
                /rpt_family="L1"
repeat_region   complement(6688. .6955)
                /rpt_family="Alu"
repeat_region   6821. .6847
                /note="(T)27"
                /rpt_type=tandem
                /rpt_unit=T
repeat_region   complement(7611. .7742)
                /rpt_family="MLTa"
repeat_region   complement(7833. .7932)
                /rpt_family="MSTc"
repeat_region   8492. .8601
                /rpt_family="Alu"
repeat_region   complement(9403. .9676)
                /rpt_family="Alu"
repeat_region   complement(10100. .10257)
                /rpt_family="Alu"
misc_feature    11124. .11167
                /note="GRAIL 2 excellent exon, frame 1"
repeat_region   11224. .11247
                /note="(ATTC)6"
                /rpt_type=tandem
                /rpt_unit=ATTC
repeat_region   12044. .12360
                /rpt_family="Alu"
repeat_region   12343. .12369
                /note="(A)27"
                /rpt_type=tandem
                /rpt_unit=A
repeat_region   complement(12372. .12811)
                /rpt_family="L1"
repeat_region   12900. .13016
                /rpt_family="Alu"
repeat_region   complement(13084. .13169)
                /rpt_family="L1"
repeat_region   complement(13597. .13889)
                /rpt_family="Alu"
repeat_region   complement(15617. .15931)
                /rpt_family="Alu"
repeat_region   15793. .15817
                /note="(T)25"
                /rpt_type=tandem
                /rpt_unit=T
repeat_region   complement(16185. .16462)
                /rpt_family="Alu"
repeat_region   complement(16604. .16902)
                /rpt_family="Alu"
repeat_region   17036. .17206
                /rpt_family="MLT1"
repeat_region   17220. .17832
                /rpt_family="Alu"
repeat_region   17626. .17649
                /note="(A)24"
                /rpt_type=tandem
                /rpt_unit=A
repeat_region   17874. .18004
                /rpt_family="MLT1f"
repeat_region   18255. .18529
                /rpt_family="Alu"
repeat_region   complement(18691. .18870)
                /rpt_family="MER3"
repeat_region   complement(18788. .18822)
                /rpt_family="MER33"
repeat_region   complement(18983. .19265)
                /rpt_family="Alu"
repeat_region   20976. .21209
                /rpt_family="Alu"
repeat_region   21298. .21414
                /rpt_family="L1"
repeat_region   complement(21413. .21577)
                /rpt_family="Alu"
repeat_region   complement(21730. .22011)
                /rpt_family="Alu"
repeat_region   22064. .22271
                /rpt_family="Alu"
misc_feature    22531. .22595
                /note="GRAIL 2 excellent exon, frame 2"
repeat_region   22780. .23054
                /rpt_family="Alu"
repeat_region   23190. .23489
                /rpt_family="Alu"
repeat_region   23849. .24002
                /rpt_family="MER39B"
repeat_region   complement(24000. .26149)
                /rpt_family="L1"
repeat_region   26149. .27398
                /rpt_family="L1"
repeat_region   27432. .27706
                /rpt_family="Alu"
repeat_region   27721. .28129
                /rpt_family="MER39"
repeat_region   28498. .28774
                /rpt_family="Alu"
repeat_region   complement(29403. .29490)
                /rpt_family="MIR2"
repeat_region   29525. .29716
                /rpt_family="Alu"
repeat_region   30196. .30275
                /rpt_family="MIR"
repeat_region   complement(30377. .30662)
                /rpt_family="Alu"
repeat_region   complement(30672. .31137)
                /rpt_family="L1"
repeat_region   31138. .31564
                /rpt_family="MER57B"
repeat_region   complement(31565. .31637)
                /rpt_family="L1"
misc_feature    join(31640. .31733,31791. .32094)
                /note="85% & 91% identity EST zb50f04.r1"
                /db_xref="dbEST:W24400"
misc_feature    31917. .32270
                /note="90% identity EST zi04c10.r1"
                /db_xref="dbEST:AA009751"
repeat_region   32354. .32655
                /rpt_family="Alu"
repeat_region   complement(32690. .32798)
                /rpt_family="Alu"
repeat_region   complement(32793. .32997)
                /rpt_family="L1"
repeat_region   33000. .33296
                /rpt_family="Alu"
repeat_region   34147. .34354
                /rpt_family="Alu"
repeat_region   complement(34383. .34822)
                /rpt_family="L1"
repeat_region   35142. .35273
                /rpt_family="MER87"
repeat_region   36021. .36304
                /rpt_family="Alu"
repeat_region   36482. .36594
                /rpt_family="MIR"
misc_feature    36830. .36949
                /note="GRAIL 2 excellent exon, frame 1"
repeat_region   complement(37800. .37874)
                /rpt_family="MER91B"
repeat_region   38633. .38910
                /rpt_family="Alu"
repeat_region   38909. .38935
```

```
                     /note="(AAT)9"
                     /rpt_type=tandem
                     /rpt_unit=AAT

  Query Match             95.5%;  Score 478.6;  DB 9;  Length 186780;
  Best Local Similarity   97.2%;  Pred. No. 2e-108;
  Matches  487;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db      25041 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACCCTTCAT 24982

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              ||||||||||||||| |||||||||||||||||||||||||||||||||||  ||||
Db      24981 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 24922

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      24921 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 24862

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24861 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 24802

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db      24801 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 24742

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      24741 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGT 24682

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24681 CTCAGCGCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 24622

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||
Db      24621 CAATGTACAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 24562

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db      24561 CATGAGTGAACTCCCATTCAC 24541


RESULT 14
AC087535/c
LOCUS       AC087535               196869 bp    DNA     linear   HTG 26-FEB-2001
DEFINITION  Homo sapiens chromosome 18 clone RP11-869L2 map 18, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC087535
VERSION     AC087535.2  GI:13123259
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 196869)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-869L2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196869)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JAN-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 25, 2001 this sequence version replaced gi:12043614.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12339
                Center clone name: 869_L_2
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 189512 bases at least Q40
                Consensus quality: 193548 bases at least Q30
                Consensus quality: 194822 bases at least Q20
                Insert size: 198000; agarose-fp
                Insert size: 195669; sum-of-contigs
                Quality coverage: 5.4 in Q20 bases; agarose-fp
                Quality coverage: 5.5 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11489: contig of 11489 bp in length
            *    11490    11589: gap of 100 bp
            *    11590    12389: contig of 800 bp in length
            *    12390    12489: gap of 100 bp
            *    12490    15301: contig of 2812 bp in length
            *    15302    15401: gap of 100 bp
            *    15402    17297: contig of 1896 bp in length
            *    17298    17397: gap of 100 bp
            *    17398    22220: contig of 4823 bp in length
            *    22221    22320: gap of 100 bp
            *    22321    27991: contig of 5671 bp in length
            *    27992    28091: gap of 100 bp
            *    28092    31242: contig of 3151 bp in length
            *    31243    31342: gap of 100 bp
            *    31343    41247: contig of 9905 bp in length
            *    41248    41347: gap of 100 bp
            *    41348    49982: contig of 8635 bp in length
            *    49983    50082: gap of 100 bp
            *    50083    64395: contig of 14313 bp in length
            *    64396    64495: gap of 100 bp
            *    64496    79369: contig of 14874 bp in length
            *    79370    79469: gap of 100 bp
            *    79470   151119: contig of 71650 bp in length
            *   151120   151219: gap of 100 bp
            *   151220   196869: contig of 45650 bp in length.
FEATURES            Location/Qualifiers
     source         1. .196869
                    /organism="Homo sapiens"
```

```
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-869L2"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .11489
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    11590. .12389
                     /note="assembly_fragment"
     misc_feature    12490. .15301
                     /note="assembly_fragment"
     misc_feature    15402. .17297
                     /note="assembly_fragment"
     misc_feature    17398. .22220
                     /note="assembly_fragment"
     misc_feature    22321. .27991
                     /note="assembly_fragment"
     misc_feature    28092. .31242
                     /note="assembly_fragment"
     misc_feature    31343. .41247
                     /note="assembly_fragment"
     misc_feature    41348. .49982
                     /note="assembly_fragment"
     misc_feature    50083. .64395
                     /note="assembly_fragment"
     misc_feature    64496. .79369
                     /note="assembly_fragment"
     misc_feature    79470. .151119
                     /note="assembly_fragment"
     misc_feature    151220. .196869
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    56352 a  38146 c  39166 g  61991 t   1214 others
ORIGIN

  Query Match             95.5%;  Score 478.6;  DB 2;  Length 196869;
  Best Local Similarity   97.2%;  Pred. No. 2e-108;
  Matches  487;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db     140922 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACACTTCAT 140863

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              ||||||||||||||| |||||||||||||||||||||||||||||||||||  ||||
Db     140862 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 140803

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     140802 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 140743

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| |||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     140742 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCCATTCAACATAGTGTTGGA 140683

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     140682 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 140623

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     140622 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGT 140563

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     140562 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 140503

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||

Db     140502 CAATGTACAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 140443

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db     140442 CATGAGTGAACTCCCATTCAC 140422


RESULT 15
AC011401/c
LOCUS       AC011401               199275 bp    DNA     linear   PRI 01-OCT-2002
DEFINITION  Homo sapiens chromosome 5 clone CTB-35K5, complete sequence.
ACCESSION   AC011401
VERSION     AC011401.9  GI:23396212
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 199275)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 199275)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 199275)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 199275)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Oct 1, 2002 this sequence version replaced gi:12830130.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.5% of Sequence;
            Estimated Total Number of Errors is 1.
FEATURES             Location/Qualifiers
     source          1. .199275
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTB-35K5"
BASE COUNT    57557 a  44170 c  42862 g  54686 t
ORIGIN

  Query Match             95.5%;  Score 478.6;  DB 9;  Length 199275;
  Best Local Similarity   97.2%;  Pred. No. 2e-108;
  Matches  487;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db     167633 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACCCTTCAT 167574

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              ||||||||||||||| |||||||||||||||||||||||||||||||||||  ||||
Db     167573 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 167514

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     167513 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 167454

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| |||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     167453 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 167394
```

```
Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     167393 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 167334

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     167333 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGT 167274

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     167273 CTCAGCGCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 167214

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||| |||||||| ||||||||||||||||||||||||||||
Db     167213 CAATGTACAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 167154

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db     167153 CATGAGTGAACTCCCATTCAC 167133
```

Search completed: November 7, 2003, 11:11:31
Job time : 1969.4 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 6, 2003, 21:13:51 ; Search time 310.835 Seconds
(without alignments)
8684.478 Million cell updates/sec

Title: US-09-939-209A-3_COPY_1_1000
Perfect score: 1000
Sequence: 1 agttcaagaccagcctgagc.....gtatgtggcagaggcatggt 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
✓	1	1000	100.0	20300	24	ABK47337	Genomic nucleotide
c	2	356.6	35.7	121600	24	ABT10748	Human breast cance
c	3	322.2	32.2	26997	22	AAS46748	Tumour suppressor
c	4	311.2	31.1	26997	22	AAS46747	Tumour suppressor
	5	296	29.6	15418	21	AAA63785	Nucleotide sequenc
	6	296	29.6	15418	24	ABS54997	Lambda clone conta
	7	296	29.6	15418	24	AAL38595	DNA of plasmid pGR
	8	296	29.6	15418	24	AAL38601	Human TERT promote
	9	293.6	29.4	27067	25	AAD50021	Human secreted pro
	10	288	28.8	30393	22	AAK67239	Human immune/haema
c	11	286.6	28.7	44147	24	ABK84481	Human cDNA differe
	12	280.8	28.1	32206	22	AAK89374	Human digestive sy
	13	270.4	27.0	20420	22	AAK73165	Human immune/haema
	14	270.4	27.0	20420	24	ABK69933	Human secreted pro
	15	267.4	26.7	19965	22	AAK73166	Human immune/haema
	16	267.4	26.7	19965	24	ABK69932	Human secreted pro
c	17	266.4	26.6	6971	24	ABL33237	Human immune syste
	18	263	26.3	165199	24	ABK83460	Human cDNA differe
c	19	263	26.3	227968	24	ABK83497	Human cDNA differe
c	20	261.4	26.1	12069	24	ABK39930	Human chemically p
c	21	260.6	26.1	20420	22	AAK73165	Human immune/haema
c	22	260.6	26.1	20420	24	ABK69933	Human secreted pro
	23	258.4	25.8	82938	24	ABV72623	Human transporter
c	24	258	25.8	19965	22	AAK73166	Human immune/haema
c	25	258	25.8	19965	24	ABK69932	Human secreted pro
	26	257.6	25.8	12069	24	ABK39930	Human chemically p
c	27	256	25.6	870	14	AAQ39248	Sequence of cosmid
c	28	253.6	25.4	80595	20	AAV83939	HC-contig derived
c	29	253	25.3	106746	21	AAA10225	Human PCTA-1 genom
c	30	252.6	25.3	80240	20	AAV83940	NC-contig derived
c	31	252.6	25.3	82938	24	ABV72623	Human transporter
c	32	250	25.0	21537	24	ABL33999	Human immune syste
	33	247.4	24.7	121600	24	ABT10748	Human breast cance
c	34	245	24.5	6286	22	AAS46591	Tumour suppressor
	35	242	24.2	1691080	24	ABX08336	Human phosphodiest
c	36	240.4	24.0	13573	24	ABL33869	Human immune syste
c	37	240.2	24.0	165199	24	ABK83460	Human cDNA differe
	38	237.4	23.7	40023	24	ABL51954	Human solute carri
c	39	235.2	23.5	21537	24	ABL33998	Human immune syste
	40	234.2	23.4	26997	22	AAS46748	Tumour suppressor
	41	234.2	23.4	80595	20	AAV83939	HC-contig derived
	42	234.2	23.4	154902	24	ABQ88198	Human osteoblast d
c	43	232.8	23.3	110000	22	AAF84800	Nucleotide sequenc
	44	232.2	23.2	80240	20	AAV83940	NC-contig derived
	45	226.8	22.7	15577	19	AAV35616	SHOX gene prelimin

ALIGNMENTS

RESULT 1
ABK47337
ID ABK47337 standard; DNA; 20300 BP.
XX
AC ABK47337;
XX
DT 18-JUN-2002 (first entry)
XX
DE Genomic nucleotide sequence encoding human RGS-4 protein.
XX
KW RGS-4; schizophrenia; human; regulator of G protein signalling 4;
KW neuroleptic; polymorphism; schizotypy; schizoid; chromosome 1q21-22;
KW gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace (4121,T)
FT /*tag= a
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4123,A)
FT /*tag= b
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4368,C)
FT /*tag= c
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4621,C)
FT /*tag= d
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4790,T)
FT /*tag= e

```
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (4816,T)
FT                   /*tag= f
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (4970,T)
FT                   /*tag= g
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace 5055,G)
FT                   /*tag= h
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (5295,A)
FT                   /*tag= i
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (5695,A)
FT                   /*tag= j
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (7375,T)
FT                   /*tag= k
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (7759,A)
FT                   /*tag= l
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (8596,A)
FT                   /*tag= m
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (9602..9610,CA)
FT                   /*tag= n
FT                   /note= "Deletion polymorphism"
FT   variation       replace (9892,A)
FT                   /*tag= o
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (9963,A)
FT                   /*tag= p
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (10132,A)
FT                   /*tag= q
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11056,C)
FT                   /*tag= r
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11091,T)
FT                   /*tag= s
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11106,A)
FT                   /*tag= t
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11774,T)
FT                   /*tag= u
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (12143,A)
FT                   /*tag= v
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (12145,T)
FT                   /*tag= w
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (14367,G)
FT                   /*tag= x
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (17027..17029,GA)
FT                   /*tag= y
FT                   /note= "Deletion polymorphism"
FT   variation       replace (17630,T)
FT                   /*tag= Z
FT                   /standard_name= "Single-nucleotide polymorphism"
XX
PN   WO200216653-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26622.
XX
PR   24-AUG-2000; 2000US-228021P.
XX
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Levitt PR, Mirnics K, Kodavali VC, Nimgaonkar VL;
XX
DR   WPI; 2002-292070/33.
XX
PT   Diagnosing, assessing susceptibility and treating schizophrenia,
PT   involves observing regulator of G-protein signalling 4, RGS4 levels in a
PT   subject -
XX
PS   Claim 1; Page 20-33; 112pp; English.
XX
CC   This invention relates to a novel method for diagnosing schizophrenia
CC   or determining susceptibility to schizophrenia in a human. The method
CC   comprises obtaining from a patient a DNA sample and detecting variations
CC   in the regulator of G-protein signalling 4 (RGS4) gene. Alternatively,
CC   the method involves measuring RGS4 mRNA or protein levels in a tissue
CC   sample from the patient and determining if there is a reduced level.
CC   The method of the invention is useful for diagnosing and determining
CC   susceptibility to schizophrenia. The invention also comprises a method
CC   that is useful for treating schizophrenia which includes a prophylactic
CC   treatment. The method of genotyping polymorphic variants in the RGS-4
CC   gene is applied to diagnosing pathologies of the schizophrenic spectrum,
CC   such as in particular schizotypy, schizoid individuals, etc. This
CC   method offers the possibility of diagnosing schizophrenia by a
CC   biological test and no longer exclusively by clinical evaluations.
CC   The present sequence represents the genomic DNA encoding the human
CC   regulator of G-protein signalling 4 (RGS4) protein used in the method of
CC   the invention. The gene for the RGS4 protein is located on human
CC   chromosome 1q21-22.
XX
SQ   Sequence 20300 BP; 6157 A; 4102 C; 3775 G; 6266 T; 0 other;

  Query Match             100.0%;  Score 1000;  DB 24;  Length 20300;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-150;
  Matches 1000;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120

Qy        121 CACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480

Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
```

```
Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780

Qy        781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840

Qy        841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900

Qy        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960

Qy        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGT 1000
              |||||||||||||||||||||||||||||||||||||||
Db        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGT 1000
```

RESULT 2
ABT10748/c
ID ABT10748 standard; cDNA; 121600 BP.
XX
AC ABT10748;
XX
DT 04-DEC-2002 (first entry)
XX
DE Human breast cancer associated coding sequence SEQ ID NO: 882.
XX
KW Human; breast specific gene; breast cancer; differential expression;
KW cytostatic; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200259271-A2.
XX
PD 01-AUG-2002.
XX
PF 25-JAN-2002; 2002WO-US02176.
XX
PR 25-JAN-2001; 2001US-263757P.
PR 25-APR-2001; 2001US-286090P.
PR 23-MAY-2001; 2001US-292517P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Orr MS, Nation M, Diggans JC, Zeng W;
XX
DR WPI; 2002-674803/72.
XX
PT Diagnosing breast cancer in a patient comprises detecting the level of
PT gene expression in cell or tissue samples, where a differential gene
PT expression is indicative of breast cancer -
XX
PS Claim 1; SEQ ID NO 882; 260pp + Sequence Listing; English.
XX
CC The present invention relates to methods of diagnosing breast cancer in a
CC patient, which comprise detecting the level of expression in a tissue
CC sample of two or more genes selected from those shown in ABT09867-
CC ABT11112, where a differential expression of the genes indicates breast
CC cancer. The methods are useful in diagnosing, treating, detecting the
CC progression, and in monitoring treatment of breast cancer in patients.
CC The methods are also useful as a screening tool for agents that modulate
CC the onset or progression of breast cancer. The breast cancer genes may be
CC used as diagnostic markers for the prediction or identification of the
CC malignant state of breast tissue, for confirming the type and progression
CC of cancer, and for drug screening and assays. The present sequence is a
CC coding sequence of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub.published_pct_sequences.
XX
SQ Sequence 121600 BP; 38372 A; 24652 C; 24048 G; 34528 T; 0 other;

Query Match 35.7%; Score 356.6; DB 24; Length 121600;
Best Local Similarity 74.7%; Pred. No. 2.7e-48;
Matches 543; Conservative 0; Mismatches 169; Indels 15; Gaps 7;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA--- 57
              || || |||||| ||||   ||||  |||||||||| |||||||||||||| |||||
Db      44776 AGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATCCAAAAAAA 44717

Qy         58 --TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
                |||| | ||| |||||       | |||||| |||||||||||  |||||| ||||||||
Db      44716 AGTTAGCCGGGCGTGGTGGCGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGG 44657

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCAC 174
              |||||  | ||||| ||  || || |||| |||| | ||| | ||||   |||| |||||
Db      44656 AGAATGGCGTGAACTTG--GGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCAC 44599

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              ||||| |  ||| || |||||||| |   ||||||| |||||||| |     | ||||||
Db      44598 TCCAGTCTGGGCGAC-AGAGCGAGACTCCATCTCAAAAAAAAAATA----TATATATATA 44544

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              || ||||||||| |||||||| | | | | | ||||||| ||| | | ||||||| | |||
Db      44543 TAGATATATATAGATATACAGATATATAGATATATATAGATATATAGATATATATAGATA 44484

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |||| ||||||| | ||||||| | ||||||| | | ||| |||| || | ||||||| |
Db      44483 TATAGATATATATAGATATATAGATATATATAGATATATAGATATATATAGATATATAGA 44424

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | ||||  ||| | | ||||| | | ||||| | | ||||| | | ||||| | | | |
Db      44423 TATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATATA 44364

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              ||||| | | ||||| | ||||||| | ||||| | ||||||| | ||||||| |||||
Db      44363 GATATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATA 44304

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || |  |||| || ||||| | | ||||| ||| | ||| ||| | | ||||||| | ||
Db      44303 TAGATGTATAGAT-ATATATAGATATATAGATATATATAGATATATAGATATATAGATAT 44245

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              ||||| | ||||| ||| | ||| ||| | ||||| | ||||| | | ||| |||||
Db      44244 ATATAGATATATAGATATATATAGATATATATATAGATATATAGATATATAGATATATAG 44185

Qy        595 ATATACACATATATACATATATACA-CATATATACATATATACACACATATATACACATA 653
              ||||| | ||||||| ||||||| |  ||||||| ||||||| | | ||||| | | |||
Db      44184 ATATATAGATATATAGATATATAGATGATATATAGATATATATAGATATATAGATATATA 44125

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              ||||| | | | ||||||||| ||| ||||| | ||||||| |  |||| |||| ||||
Db      44124 GATATATAGATATATAGATATATATAGATATATAGATATATAGATATATAGATATATATA 44065

Qy        714 TATATAT 720
              |||||||
Db      44064 GATATAT 44058
```

RESULT 3

```
AAS46748/c
ID   AAS46748 standard; DNA; 26997 BP.
XX
AC   AAS46748;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #472.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW   cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02955.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumour suppressor
PT   genes and oncogenes, useful in designing primers and probes for
PT   analysing diseases associated with cytosine methylation state e.g.
PT   cancer  -
XX
PS   Claim 1; SEQ ID No 472; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and
CC   oncogenes having a sequence taken from 536 (actually 533 since
CC   numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC   (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC   peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC   form part of a set of probes for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms and also to be used in an
CC   array for analysing diseases associated with CpG dinucleotides e.g.
CC   cancers and tumours. The probes can also be used in a method for
CC   ascertaining genetic and/or epigenetic parameters for the diagnosis
CC   and/or therapy of existing diseases or the predisposition to specific
CC   diseases, by analysing cytosine methylations. The parameters may be
CC   compared to another set of genetic and/or epigenetic parameters, the
CC   differences serving as basis for diagnosis and/or prognosis events which
CC   are disadvantageous to patients. The present sequence is one of the
CC   533 genomic sequences derived from tumour suppressor genes and
CC   oncogenes. Sequences with even numbered Seq ID numbers are the
CC   complementary sequence of the corresponding odd numbered sequence (e.g.
CC   ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC   is missing).
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 26997 BP; 7650 A; 354 C; 5924 G; 13069 T; 0 other;

  Query Match             32.2%;  Score 322.2;  DB 22;  Length 26997;
  Best Local Similarity   66.0%;  Pred. No. 8.2e-43;
  Matches  498;  Conservative    0;  Mismatches  253;  Indels    4;  Gaps    2;

Qy        207 TCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACA 266
              | |  |||||||| |      | ||||||||||||||||||||||||||||||| | | | |
Db       6221 TAAATAAAAAAAATATAATATATATATATATATATATATATATACACACATATATATA 6162

Qy        267 TATATATATACACATATATATACACATATATACATATATACACATATATACACATATATA 326
              | |||||||||||| | ||||||| ||||| ||| ||||| ||||||||||||||||||
Db       6161 AACATATATACACACACATATACATATATACACACATATATACATATATACACATATATA 6102

Qy        327 TACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATA 386
              || | ||| | || || | | ||||||| | | |  |||| | ||||||||| |||||||
Db       6101 TATATATATACATATATATACATATATATATACACATATATATATATATACATATATATA 6042

Qy        387 CACACATATATACACATATATACACACATATATACACATATATACACATATATACACATA 446
              || | | |||||||||||||||| | | ||| ||| | ||| ||| ||||||| ||| |||
Db       6041 CATATACATATATACACATATATATATATATACATATATATACATATACATATACACATATA 5982

Qy        447 TACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATA 505
              || | ||||| | | | ||| | ||||||||||||||||||| ||||||| | | ||| |
Db       5981 TATATATATACATATACATACATATATATACACATATATATATATATACATATACATACA 5922

Qy        506 TACACATATATACACACATATATACACATATATACACATATATATACACATATATACACA 565
              || | ||| | | | | ||||||| |||||||||||| |||||| | | ||| | ||| |
Db       5921 TATATATACACATATATATATATATACATATATACACGTATATACATATATACACACATA 5862

Qy        566 TATATACATATATACACATATATATACATATATACACATATATACATATATACACA---T 622
              | |||| |||  || | | ||||||| |||||||| ||||||| ||||||| |||   |
Db       5861 TTTATATATACGTATATACATATATATATATATACATATATATATATATATACATATT 5802

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
              ||||| ||||||| | |||||||||| | ||||||||| | | | ||| | ||| ||| ||
Db       5801 ATATATATATATATATACATATATAAATATACATATATATATATATATACATATATACAT 5742

Qy        683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742
              ||| | |||||||| | |||||||||| |||| |||||           || |      ||
Db       5741 ATATATATATATAAATATATATATATATATACGTATATATATATACACACACACCATA 5682

Qy        743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802
                  ||   | |     | | |  |||| |   |||   |  | |  | |   | |
Db       5681 AAATACTATACACCCATAAAAAAAAAAAAAAATAACATTTACAACAACTTAAATAAAATTA 5622

Qy        803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862
                |   | |   | |  | ||  | |     | | | |   | | | ||| |  ||
Db       5621 AAAACCATTATTCTAACCGAAATAACTCAAAAATAAAAAACCAAATATTATATATTCTCA 5562

Qy        863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTA 922
              ||   | ||   | ||  ||  | | | | | | ||| | | |   ||||| |
Db       5561 CTTATAACTAAAAACTAAACTATAAAAACACAAAAACATAAAAATAATACAATAAACTTT 5502

Qy        923 AAGGATGGGCCAAGGATTTAAATGTTATAGAAGAA 957
              ||  |     ||  |   ||  || | || ||
Db       5501 AAAAACTCTAAAAAAAACATAAAAATAAATAAAAA 5467


RESULT 4
AAS46747/c
ID   AAS46747 standard; DNA; 26997 BP.
XX
AC   AAS46747;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #471.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW   cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
```

```
PF   15-MAR-2001; 2001WO-EP02955.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumour suppressor
PT   genes and oncogenes, useful in designing primers and probes for
PT   analysing diseases associated with cytosine methylation state e.g.
PT   cancer -
XX
PS   Claim 1; SEQ ID No 471; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and
CC   oncogenes having a sequence taken from 536 (actually 533 since
CC   numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC   (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC   peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC   form part of a set of probes for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms and also to be used in an
CC   array for analysing diseases associated with CpG dinucleotides e.g.
CC   cancers and tumours. The probes can also be used in a method for
CC   ascertaining genetic and/or epigenetic parameters for the diagnosis
CC   and/or therapy of existing diseases or the predisposition to specific
CC   diseases, by analysing cytosine methylations. The parameters may be
CC   compared to another set of genetic and/or epigenetic parameters, the
CC   differences serving as basis for diagnosis and/or prognosis events which
CC   are disadvantageous to patients. The present sequence is one of the
CC   533 genomic sequences derived from tumour suppressor genes and
CC   oncogenes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 26997 BP; 7233 A; 354 C; 6190 G; 13220 T; 0 other;

  Query Match             31.1%;  Score 311.2;  DB 22;  Length 26997;
  Best Local Similarity   76.9%;  Pred. No. 4.5e-41;
  Matches  409;  Conservative    0;  Mismatches  113;  Indels   10;  Gaps    2;

Qy        203 CATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACA 262
              | | ||| ||| | | ||       | ||||||||||||||||||| | | || || | | |
Db      21339 CTTTTCATAAATACATAATATTCCATAATATATATATATATATACACACGCATATATA 21280

Qy        263 CACATATATATATACACATATATATACACATATATACATATATACACATATATACACATA 322
              ||||||||    | ||| ||| ||||| | |||||||||||||| ||| ||||||| |  ||
Db      21279 CACATATACCCACACATATACATATATATATATATACATATACACATATATATATATCTA 21220

Qy        323 TATATACACATACATATGT-----ACACATATATATACACATATGTATACACATATATAC 377
               |||||||| | |||  |     | | ||| | ||| |||||| |||| ||| ||||||
Db      21219 CATATACACACATATACATATATAATATATACACATATACATATATATATACACATATAC 21160

Qy        378 ACATATATACACACATATATACACATATATACACACATATATACACATATATACACATAT 437
              ||||||| | || || ||||| |||||||||| | ||||||||||||| |||| ||| |||
Db      21159 ACATATACATACGCACATATAAACATATATATATACATATATACACGTATACACATATAC 21100

Qy        438 ATACACATATACACATATATACACATATATACATATATACACATATATATA-----ATAT 492
              ||| ||| ||| | ||||||||| ||||| ||||||| | | ||| |||||     |||
Db      21099 ATATACACATATATATATATACATATATACACATATACATATATACATATATATATATAC 21040

Qy        493 ACACACATATATATACACATATATACACACATATATACACATATATACACATATATATAC 552
              | | | | | ||||||| ||||| | | | ||||||| ||||||||||||||||| |||
```

```
Db      21039 ATATATACACATATACATATATACATATATATATATATACATATATACACATATACATAT 20980

Qy        553 ACATATATACACATATATACATATATACACATATATATACATATATACACATATATACAT 612
              | | ||||| | |||||||||||||||||||||||| ||| ||| ||| | |||||||| |
Db      20979 ATACATATATATATATATACATATATACACATATACATATATACATATATATATATATAC 20920

Qy        613 ATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGAT 672
              ||||| ||| |||||||||||||| | ||||||| | ||| ||||| | | | ||| ||
Db      20919 ATATATACACATATACATATATACATATATATATATATATATATATATATATATACAT 20860

Qy        673 ATACATATATATACACATATATATACGTATATATATGTATATATATATGCTC 724
              ||| |||||||||| |||||| ||  ||||||||| |||||||||| | |
Db      20859 ATATATATATATACATATATATCTATATATATATATATATATATATATACAC 20808


RESULT 5
AAA63785
ID   AAA63785 standard; DNA; 15418 BP.
XX
AC   AAA63785;
XX
DT   04-DEC-2000  (first entry)
XX
DE   Nucleotide sequence of the human TERT promoter and gene from pGRN142.
XX
KW   Telomerase reverse transcriptase; TERT; promoter; telomerase complex;
KW   telomere length; hyperplasic disease; cancer; oncolytic virus;
KW   cis-acting transcriptional control sequence; viral replication;
KW   cell proliferation; aging; immunological disorder; infertility; ss.
XX
OS   Homo sapiens.
XX
PN   WO200046355-A2.
XX
PD   10-AUG-2000.
XX
PF   04-FEB-2000; 2000WO-US03104.
XX
PR   04-FEB-1999;   99US-0244438.
XX
PA   (GERO-) GERON CORP.
XX
PI   Morin GB,  Lichtsteiner S,  Vasserot A,  Adams R,  Cardoza LM;
PI   Lebkowski JS;
XX
DR   WPI; 2000-532898/48.
XX
PT   New polynucleotides comprising cis-acting transcriptional control
PT   sequences, e.g. promoter sequence, of telomerase reverse transcriptase
PT   genes, useful in the treatment of cancer -
XX
PS   Claim 9; Page 58-61; 63pp; English.
XX
CC   The present sequence represents the human telomerase reverse
CC   transcriptase (TERT) promoter and gene. TERT is part of the telomerase
CC   complex responsible for maintaining telomere length and increasing
CC   replicative capacity of progenitor cells. Telomerase activity is
CC   turned off in mature differentiated cells, but is turned back on again
CC   in hyperplasic diseases, including many cancers. The polynucleotide
CC   comprises cis-acting transcriptional control sequences, e.g. promoter
CC   sequence. These promoter sequences are used to produce oncolytic viruses,
CC   in which a toxin or genetic element essential for viral replication is
CC   placed under the control of a TERT promoter. As a result, the virus
CC   replicates preferentially in cells expressing TERT, and selectively
CC   lyses cancer cells. The oncolytic viruses are useful for treating cancer
CC   in humans or animals. The TERT promoter sequences are useful in the
CC   treatment of cancer and other diseases of cell proliferation such as
CC   degenerative and aging processes and diseases of aging, immunological
CC   disorders, or infertility.
XX
SQ   Sequence 15418 BP; 4518 A; 3797 C; 3765 G; 3338 T; 0 other;

  Query Match             29.6%;  Score 296;  DB 21;  Length 15418;
```

```
  Best Local Similarity   69.2%;  Pred. No. 1.2e-38;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||    ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||      |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||||  || | | ||| ||| | | |||||||  | ||| |||||  ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | |||||||  |||||  | |||||  | |  |||  ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||  |||||  | |  |||||  |||  |||   ||||  |||  |||||||||  |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              ||||  |    |||||   ||  | |  |||||  |  |||  |||  |||  |  |  ||  |  |    |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |  |  || ||  |    | | |  |  | |     ||  | || |||  | |  ||||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |        || | |||  | |   |||  |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 6
ABS54997
ID   ABS54997 standard; DNA; 15418 BP.
XX
AC   ABS54997;
XX
DT   10-DEC-2002  (first entry)
XX
DE   Lambda clone containing human TERT genomic insert.
XX
KW   Telomerase reverse transcriptase; TERT; replication-conditional virus;
KW   adenovirus replication gene; cancer cell; lung; pancreatic cancer;
KW   medulloblastoma; cervical carcinoma; fibrosarcoma; osteosarcoma;
KW   cytolysis; replication defective adenovirus vector; congenital defect;
KW   proinflammatory; antiinflammatory; heterologous effector gene;
KW   cancer therapy; cytostatic; gene therapy; lambda clone; human; ds.
XX
OS   Bacteriophage lambda.
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..43
FT                   /*tag=  a
FT                   /note= "From Lambda clone"
FT   misc_feature    44..15375
FT                   /*tag=  b
FT                   /note= "Human telomerase reverse transcriptase (hTERT)
FT                          sequence including the hTERT promoter and
FT                          upstream sequence"
FT   misc_feature    15376..15418
FT                   /*tag=  c
FT                   /note= "From Lambda clone"
XX
PN   WO200253760-A2.
XX
PD   11-JUL-2002.
XX
PF   17-DEC-2001; 2001WO-US48785.
XX
PR   18-DEC-2000; 2000US-256418P.
XX
PA   (GERO-) GERON CORP.
XX
PI   Irving JM,  Lebkowski JS;
XX
DR   WPI; 2002-723123/78.
XX
PT   Novel replication-conditional virus useful for cytolysis of target
PT   cells e.g. cancer cells and preparing a medicament for treating cancer,
PT   comprises heterologous replication element in an adenovirus-based
PT   construct  -
XX
PS   Claim 11; Page 26-29; 32pp; English.
XX
CC   The present invention relates to a new replication-conditional virus with
CC   a genome comprising adenovirus replication genes and one or more
CC   heterologous gene(s) that functionally replaces one or more adenovirus
CC   gene(s) required for replication or assembly of the virus. The invention
CC   is useful for killing a cancer cell (such as lung, pancreatic cancer,
CC   medulloblastoma, cervical carcinoma, fibrosarcoma or osteosarcoma),
CC   killing a cell expressing TERT (telomerase reverse transcriptase), and in
CC   preparing a medicament for treating cancer and a condition associated
CC   with increased expression of TERT in affected cells, in a subject.
CC   The invention is also useful for cytolysis of specific target cells.
CC   The invention is further useful for producing replication defective
CC   adenovirus vector which is useful for transient expression of a
CC   heterologous therapeutic gene to correct a congenital defect, introducing
CC   proinflammatory or antiinflammatory activity, enhancing telomerase
CC   function, and delivering heterologous effector genes that induce killing
CC   of the transduced cells. The invention is more safe for use in cancer
CC   therapy. The present nucleic acid sequence represents the human TERT
CC   sequence contained within a lambda clone sequence of the invention.
XX
SQ   Sequence 15418 BP; 4518 A; 3797 C; 3765 G; 3338 T; 0 other;

  Query Match           29.6%;  Score 296;  DB 24;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 1.2e-38;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||    ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||      |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
```

```
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              |||||||||| ||||| ||| |||||| ||  | |||||||| || ||||      ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | ||||||||| ||||| | | ||||| ||| |||  |||| ||| |||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |  |   || || |       | | | | ||     ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |         || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 7
AAL38595
ID AAL38595 standard; DNA; 15418 BP.
XX
AC AAL38595;
XX
DT 16-AUG-2002 (first entry)
XX
DE DNA of plasmid pGRN144 containing human hTERT gene.
XX
KW Anticonvulsant; cerebroprotective; vasotropic; nootropic; stem cell;
KW neuroprotective; antiparkinsonian; antiinflammatory; undifferentiated;
KW cardiant; transcriptional control element; human embryonic; tissue graft;
KW regenerative medicine; tissue reconstitution; neural progenitor cell;
KW nerve system; epilepsy; stroke; ischaemia; Huntington's disease;
KW Parkinson's disease; multiple sclerosis; leukodystrophy; neuritis;
KW hepatocyte; liver damage; cardiomyocyte; cryoinjury; cardiac muscle;
KW pGRN144; human hTERT gene; chimeric; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
XX
PN WO200242445-A2.
XX
PD 30-MAY-2002.
XX
PF 26-NOV-2001; 2001WO-US44309.
XX
PR 27-NOV-2000; 2000US-253357P.
PR 27-NOV-2000; 2000US-253443P.
PR 13-FEB-2001; 2001US-0783203.
XX
PA (GERO-) GERON CORP.
XX
PI Gold JD, Lebrowski JS;
XX
DR WPI; 2002-479952/51.
XX
PT Depleting a cell e.g., human embryonic stem cell population of
PT undifferentiated stem cells (UC) for use in regenerative medicine
PT comprises genetically altering UC in a population to express nucleic
PT acid encoding a lethal product -
XX
PS Example 9; Page 62-66; 67pp; English.
XX
CC The invention relates to a system for depleting a cell population of
CC undifferentiated stem cells, by introducing nucleic acids of structure
CC P-X, where X is a nucleic acid encoding product that is lethal to a cell
CC in which it is expressed; or renders a cell in which it is expressed
CC susceptible to the lethal effect of an external agent, and P is a
CC transcriptional control element causing X to be preferentially expressed
CC in the undifferentiated stem cells. The system is used for depleting a
CC cell population (preferably, human embryonic stem cells) of
CC undifferentiated stem cells. A population of differentiated cells is
CC useful in regenerative medicine, and for preparing antibodies and cDNA
CC libraries that are specific for a differentiated phenotype. The cell
CC populations are also useful for drug screening and therapeutic
CC applications. The differentiated cells are useful for tissue
CC reconstitution or regeneration in a human patient in need of treatment.
CC The cells are administered in a manner that permits to graft to the
CC intended tissue side and reconstitute or regenerate the functionally
CC deficient area. The neural progenitor cells are useful for treating acute
CC or chronic damage to the nerve system e.g. epilepsy, stroke, ischaemia,
CC Huntington's disease, Parkinson's disease, multiple sclerosis,
CC leukodystrophies, neuritis, etc. The hepatocytes and hepatocyte
CC precursors are useful for assessing animal models for ability to repair
CC liver damage. The cardiomyocyte population is useful for assessing animal
CC models for cryoinjury, regenerating cardiac muscle and to treat
CC insufficient cardiac function. This polynucleotide sequence represents
CC the DNA of plasmid pGRN144 containing the human hTERT gene relating to
CC the invention.
XX
SQ Sequence 15418 BP; 4519 A; 3797 C; 3764 G; 3338 T; 0 other;

Query Match 29.6%; Score 296; DB 24; Length 15418;
Best Local Similarity 69.2%; Pred. No. 1.2e-38;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| ||||||||||  ||| || | ||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | ||  |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||    | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              |||||||||| ||||| ||| ||||| ||  | |||||||| || ||||      ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
```

```
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593

Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653

Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685

Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 8
AAL38601
ID   AAL38601 standard; DNA; 15418 BP.
XX
AC   AAL38601;
XX
DT   16-AUG-2002  (first entry)
XX
DE   Human TERT promoter and upstream sequence.
XX
KW   Cytostatic; glycosyltransferase; tumour; cell-surface carbohydrate;
KW   tissue specific; transcriptional control element; cancer cell;
KW   gene therapy; human TERT; telomerase reverse transcriptase; ds.
XX
OS   Homo sapiens.
XX
PN   WO200242468-A2.
XX
PD   30-MAY-2002.
XX
PF   26-NOV-2001; 2001WO-US44306.
XX
PR   27-NOV-2000; 2000US-253395P.
XX
PA   (GERO-) GERON CORP.
XX
PI   Schiff MJ;
XX
DR   WPI; 2002-479954/51.
XX
PT   New polynucleotide encoding glycosyltransferase enzymes including histo
PT   blood group transferase useful for treating conditions associated with
PT   hyperproliferation, such as cancers and other neoplasias  -
XX
PS   Claim 9; Page 17-20; 49pp; English.
XX
CC   The invention relates to a polynucleotide comprising an encoding sequence
CC   for a glycosyltransferase, under control of a heterologous tumour
CC   specific or tissue specific transcriptional control element, where
CC   expression of the polynucleotide in a human cell causes the cell to
CC   express a cell-surface carbohydrate determinant to which some or all
CC   humans have a naturally occurring antibody. The polynucleotide of the
CC   invention is useful for killing cancer cells and for preparing a
CC   medicament for the treatment of cancer. The polynucleotide sequence of
CC   the invention can be used to treat disorders by gene therapy. This
CC   polynucleotide sequence represents the human TERT (telomerase reverse
CC   transcriptase) promoter and upstream sequence related to the invention.
XX
SQ   Sequence 15418 BP; 4518 A; 3797 C; 3765 G; 3338 T; 0 other;

  Query Match             29.6%;  Score 296;  DB 24;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 1.2e-38;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59

Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119

Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179

Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239

Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299

Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354

Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414

Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474

Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533

Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593

Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653

Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685

Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 9
AAD50021
ID   AAD50021 standard; DNA; 27067 BP.
XX
AC   AAD50021;
XX
DT   24-MAR-2003  (first entry)
XX
DE   Human secreted protein encoding gene.
XX
KW   Human; secreted protein; cancer; gene therapy; cytostatic; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace (921, T)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (1781, T)
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (1850, A)
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   CDS             2031..25941
FT                   /*tag=  d
FT                   /product= "Human secreted protein"
```

```
FT   exon            2031..2094
FT                   /*tag= e
FT   intron          2095..5569
FT                   /*tag= f
FT   variation       replace (2839, G)
FT                   /*tag= g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (3730, A)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            5570..5762
FT                   /*tag= i
FT   intron          5763..7571
FT                   /*tag= j
FT   variation       replace (6631, A)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (6945, A)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (6952, T)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (7457, A)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            7572..7741
FT                   /*tag= o
FT   intron          7742..10000
FT                   /*tag= p
FT   variation       replace (7830, A)
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (8089, C)
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (8551, T)
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (9269, C)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (9362, T)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (9782, T)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            10001..10173
FT                   /*tag= w
FT   intron          10174..10298
FT                   /*tag= x
FT   exon            10299..10485
FT                   /*tag= y
FT   intron          10486..12027
FT                   /*tag= z
FT   variation       replace (11493, A, T)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            12028..12193
FT                   /*tag= ab
FT   intron          12194..25821
FT                   /*tag= ac
FT   variation       replace (12260, G)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (13086, C)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (13183, C)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (21240, G)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (21695, G)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (22058, T)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (22233, A)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (22245, -)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (22375, T)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (23042, -, T)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (23344, C)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (23873, -)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (24764, T)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (23873, -)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (24764, T)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (24939, C)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (24945, A)
FT                   /*tag= at
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (25092, T)
FT                   /*tag= au
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (25428, G)
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (25513, T)
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (25684, T)
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            25822..25941
FT                   /*tag= ay
FT   variation       replace (26165, -)
FT                   /*tag= az
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "These features corresponds to secreted
FT                   protein #1 (AAE32429) encoding genomic region"
FT   variation       replace (1022, T)
FT                   /*tag= ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (1882, T)
FT                   /*tag= bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (1951, A)
FT                   /*tag= bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            2132..2195
FT                   /*tag= bd
FT   intron          2196..5670
FT                   /*tag= be
```

```
FT   variation       replace (2940, G)
FT                   /*tag=  bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (3831, A)
FT                   /*tag=  bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            5671..5863
FT                   /*tag=  bh
FT   intron          5864..7672
FT                   /*tag=  bi
FT   variation       replace (6732, A)
FT                   /*tag=  bj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (7558, A)
FT                   /*tag=  bk
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            7673..7761
FT                   /*tag=  bl
FT   intron          7762..9149
FT                   /*tag=  bm
FT   variation       replace (7931, A)
FT                   /*tag=  bn
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (8190, C)
FT                   /*tag=  bo
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace (8652, T)
FT                   /*tag=  bp
FT                   /standard_name= "Single nucleotide polymorphism"

  Query Match             29.4%;  Score 293.6;  DB 25;  Length 27067;
  Best Local Similarity   75.8%;  Pred. No. 2.7e-38;
  Matches  391;  Conservative    0;  Mismatches  119;  Indels    6;  Gaps    2;

Qy        230 ATATATATATATATATATATATACACACACACACACATATAT-ATATACACATATATATA 288
              || ||||  |  | | |  ||| |   | ||| |||||| | ||| | ||||| | |
Db      26509 ATGTATAACTTAAAAATTTCATATATGTGTATACATATATATGAAATATATATATACAGA 26568

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |||| ||||| || |||||| |||||||||||||||||||  | | | ||||  |||||||
Db      26569 CACACATATATATGTATACATATATATACACATATATATGTATACATATATACACACATA 26628

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| || |||||| |||||||| |||||||||||||||||||||| ||| | | ||| | |
Db      26629 TATGTATACATATATATACACACATATACACATATATACACACACATACATATATACACA 26688

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| ||||| ||||||||||||||| ||  |||||||| |  |||| | | ||| | |
Db      26689 CACATATATACACACATATATACACACATGCACACATATATATGTATACATATATACACA 26748

Qy        469 CATATATACACATATATATA-----ATATACACACATATATATACACATATATACACACA 523
              ||| |||||  |||||| |     |||||||||||||||||||||| ||||||||||
Db      26749 CATGTATACGTATATATACACACATATATACACACATATATATACACACATATACACACA 26808

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              || | ||||||||||||||| |||||||||| ||||| ||| | ||| |||| || |||
Db      26809 TACACACACATATATACACACATATATACACACATATATACACACATATATATGTATACA 26868

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||||||||| | ||| | | | ||||| | ||||| ||||| ||||||| ||| | | |
Db      26869 TATATATACACACATATATACACATACACACATACATATATACACATATACACATATACA 26928

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
               | ||| ||||| | |   | ||||| ||||||| | || || |||||| |||||  | |
Db      26929 CACACATATACACACATGTATACATATATATACACACATGTATACATATGTATACACACA 26988

Qy        704 TATATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              |||| | |||| ||||    |      | |||| | |
Db      26989 CATATATGTATACATATATACACACATACATATGTG 27024


RESULT 10
AAK67239
```

```
ID   AAK67239 standard; DNA; 30393 BP.
XX
AC   AAK67239;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22051.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
```

PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 22051; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 30393 BP; 8271 A; 7315 C; 8076 G; 6731 T; 0 other;

Query Match 28.8%; Score 288; DB 22; Length 30393;
Best Local Similarity 76.1%; Pred. No. 2.1e-37;
Matches 382; Conservative 0; Mismatches 115; Indels 5; Gaps 2;

```
Qy      227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
            | ||||||||| | ||||| ||||||| | | |||||||||| | | | | ||| |||||||
Db    14755 TATATATATACACATATACATATATATATAGACACACACACACACACACACATATATATA 14814

Qy      287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
            ||||| | ||||||||||||||||| ||||||| ||||| | ||| |||||||| |||| |
Db    14815 CACACACACATACATATATACACACATATACATATATACACACATATACATATATACATA 14874

Qy      347 TATA----TATACACATATGTATACACATATATACACATATATACACACATATATACACA 402
            ||||    | |||| ||||    | | | ||||| ||| ||||||| | ||||| |||||
Db    14875 TATACACGTGTACATATATACGCATATACATATATACACATATACATATATATACACACA 14934

Qy      403 TATATACACACATATATACACATATATACACATATATACACATATACACATATATACACA 462
            |||| | | | | ||||| | | | ||  |||||||||| |||||||| ||||||||||
Db    14935 TATACATATATACATATATATACACATGTACATATATACGCATATACATATATATACACA 14994

Qy      463 TATATACATATATACACATATATATA-ATATACACACATATATATACACATATATACACA 521
            |||| | |||||||||||  |||| | ||||||| | ||| | ||  |||||||||| ||
Db    14995 TATACATATATATACACACCTATACATATATACATATATACACATGTACATATATACGCA 15054

Qy      522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
            ||| ||| | | | | | | || |||||| |||||| |||||||| |  |||| | |
Db    15055 TATACATATATACACACATATACATGTATACATATATACACACATATACATGTATACATA 15114

Qy      582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
            ||| | ||  | ||||| ||  |||||||||||||||||||||| | ||||| ||||| |
Db    15115 TATACACATGTACATATATACGCATATACATATATACACATATACATATATACACACATA 15174
```

```
Qy        642 TATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTA 701
              ||||| ||||| ||| |||   ||||| |||  |||||| ||| | | | |||||| ||
Db      15175 CATATATACATATATACACATGTACATATATACGCATATACATATATACACATATACATA 15234

Qy        702 TATATATGTATATATATATGCT 723
              |||| |  | || |||||| ||
Db      15235 TATACACATGTACATATATACT 15256
```

RESULT 11
ABK84481/c
ID ABK84481 standard; cDNA; 44147 BP.
XX
AC ABK84481;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #1052.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US30821.
XX
PR 03-OCT-2000; 2000US-237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity -
XX
PS Claim 1; SEQ ID No 1052; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection,
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 44147 BP; 9813 A; 11709 C; 11789 G; 10833 T; 3 other;

Query Match 28.7%; Score 286.6; DB 24; Length 44147;
Best Local Similarity 72.9%; Pred. No. 3.4e-37;
Matches 453; Conservative 0; Mismatches 154; Indels 14; Gaps 6;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57
              ||||  |||||||||||| |||||||| |||||||| |||||| |||||||||    ||||
Db      10942 AGTTTGAGACCAGCCTGACCAACATGGAGAAACCCCGTCTCTATTAAAAATACAAAAAAA 10883

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              |||| ||||| |||||  ||| |||||||||||||||||||| | |||| || |||||||
Db      10882 TTAGCCAGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTTGGAAGGCTGAAGCAGGAG 10823

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTC 176
              || | |||||||||  | || || ||||||| ||| ||| | ||||||||||||||||||
Db      10822 AAGCGCTTGAACC--CGGGAGGCGGAGGTTTCGGTGAGCCCAGATCATGCCATTGCACTG 10765

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATA 236
              |||||  |||||||||||| ||||      || |   || ||| |   |||| |||||||
Db      10764 CAGCCTGGGCAACAAGAGCAAAAT------TCCGTTTAATAAATAATTATTTAATATATA 10711

Qy        237 TATATATATATATATACACACACACACACATATATATATACACATATATATACACATATA 296
              ||||||| | | | || | | | | |||| ||||||||||  ||||||||||   ||
Db      10710 TATATATGTGTGTGTATATATATATACACGTATATATATACGTATATATATACACGCGTA 10651

Qy        297 TACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACA 356
              || |||||   | ||||||||||  |||||| || || |||| ||| | |||||||| |
Db      10650 TATATATACGTATATATATACACGCGTATATATACGTATATATATACGCGTATATATATA 10591

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
              | ||| |||| ||| || | | | |||| | | |||  || | ||||| ||  |||||
Db      10590 CGTATATATATACACGTACATATACGTATATATATATACGTATATATATACACGTACATA 10531

Qy        417 TATACACATATATACACATATATACACATATACACATATATACACATATATACATATATA 476
              ||      ||||||| ||| ||||||  ||||| | | | ||| | |  |||| ||| | |
Db      10530 TACGTGTATATATATACACATATACGTATATATATACACATATATACGTATATATACACA 10471

Qy        477 CACATATATATAATATACACACATATATATACACATATATACACACATATATACACATAT 536
              | ||||| |  ||||  | |||| |||||||| |||||||  |   ||   |  || |||
Db      10470 TATATATACA-CATATGTATACATCTATATACATATATATAAAATAATTATTTTACGTAT 10412

Qy        537 ATACACATATATATACACATATATA-CACATATATACATATATACACATATATATACATA 595
              |||          |   | |||  ||  | |  ||||| | |||| | |||||||||| |||
Db      10411 ATAATTTACGTAAATAATATACGTATAATAATTATACGTGTATATATATATATATATATA 10352

Qy        596 TATACACATATATACATATAT 616
              | ||||| ||||| ||||||
Db      10351 CACACACACATATATATATAT 10331
```

RESULT 12
AAK89374
ID AAK89374 standard; DNA; 32206 BP.
XX
AC AAK89374;
XX

DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 2950.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01324.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.

```
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-502630/55.
XX
PT   Polynucleotides encoding digestive system antigens, useful for
PT   diagnosing, treating, preventing and/or prognozing disorders of the
PT   digestive system, particularly cancer and cancer metastases -
XX
PS   Disclosure; SEQ ID NO 2950; 986pp; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human digestive system antigens. These can be used in the
CC   diagnosis, treatment and prevention of digestive system disorders,
CC   including cancer, Meckel's diverticulum, bacterial or parasitic
CC   infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC   ulcerative colitis. The present sequence is a genomic DNA fragment
CC   encoding a digestive system antigen of the invention.
XX
SQ   Sequence 32206 BP; 9503 A; 5480 C; 7017 G; 10206 T; 0 other;

  Query Match             28.1%;  Score 280.8;  DB 22;  Length 32206;
  Best Local Similarity   74.4%;  Pred. No. 2.9e-36;
  Matches  369;  Conservative    0;  Mismatches  122;  Indels    5;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              ||||| |||||||| | ||| | | | | ||| | |  |||||||| ||| | | |||||
Db      21803 TATATGTATATATACACATACACATATATACATATATGTATATATACACATACACATATA 21862

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||||||||  |||||||||||||| | | |  ||| |||||| | | ||||  ||| ||
Db      21863 TACATATATGTATATATACACATACACATATGTATGTATACATACATATATACACATATG 21922

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||||||  |||||  |||| ||| || ||  | ||||| |  || ||||||||| | | ||
Db      21923 TATATATACATACATATATACACATGTGTATATATACATGCAGATATATACACGTGTGTA 21982

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | ||| ||||  | ||| || |||||||| |||| | || | ||| | | |||  ||||
Db      21983 TATACACATATGTATATACATGCACATATACACACGTGTATATATACACATACACGTATA 22042

Qy        469 CATATATACACATATATATA-----ATATACACACATATATATACACATATATACACACA 523
              ||| ||| ||| |||||||      | ||| |||||| |||||| ||||| | | | |||
Db      22043 TATACATATACACATATATACATACACATATACACATGTATATATATACATACACATATACA 22102

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              || || | ||| ||||| ||||| |||||  |||||| ||||| ||| ||| |||||
Db      22103 CATGTATATATACATACATATATACATACATGTATATATACATACATATATACATACATG 22162

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||||||||||| ||| | | ||| |  |||||| ||||| ||||||||||||||  |||
Db      22163 TATATATACATACATATATACATACATGTATATATACATACATACATATATACACGTATA 22222

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              |||||| | | ||||||||||||||| ||| | | |||||| || | || ||||  || |
Db      22223 TATACATACATATATACACACACATATATACATACATATATGCATACATGTATATATACA 22282

Qy        704 TATATGTATATATATA 719
              || |  |||||| |||
Db      22283 TACAGATATATACATA 22298


RESULT 13
AAK73165
ID   AAK73165 standard; DNA; 20420 BP.
XX
AC   AAK73165;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27977.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
```

PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 27977; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 20420 BP; 4834 A; 5403 C; 5630 G; 4553 T; 0 other;

Query Match 27.0%; Score 270.4; DB 22; Length 20420;
Best Local Similarity 61.4%; Pred. No. 1.3e-34;
Matches 467; Conservative 0; Mismatches 291; Indels 2; Gaps 2;

```
Qy       144 GGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTC 203
            | | | | | || |  || ||  ||   |||| |||||      |        |
Db      4055 GCTGGAGTGCAGTAGCATGATCACAGCTCACTGCAGCCATATATATGTATATATATATAT 4114

Qy       204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
             || | | |||| | | |       | |||||||  | | |||||||||| | | | | |
Db      4115 ATATGAAAAAATATATATATATATATATATATGAAAAAATATATATAAATATATATATGA 4174

Qy       264 ACATATATATATACACATATATATACACATATATACATATATACACAT-ATATACACATA 322
             | |||||||||||| | ||||||||  | |||||||  | ||||| | || ||||| | | |
Db      4175 AAATATATATATAAATATATATATGAAAATATATATAAATATATATATAATATATATAAA 4234

Qy       323 TATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATA 382
             |||||| | ||| | || || | ||  | ||| | |||  |||| | ||| ||| | |||
Db      4235 TATATATATATATAAATATATATATGAAAATATATATAAATATATATATAAATATATATA 4294

Qy       383 TATACACACATATATACACATATATACACACATATATACACATATATACACATATATACA 442
             |  | | | |||||| | |||||||   | |||||||  | | |||||  | ||||  | |
Db      4295 TGAAAATATATATATGAAAATATATATGAAAATATATATATAAATATATATATATGAAAA 4354

Qy       443 CATATACACATATATACACATATATACATATATACACATATATATAATATACACACATAT 502
             ||||| | | ||||| | ||  | | |||||||| | |||||||| ||||  | | ||||
Db      4355 TATATATATAAATATATATATGAAAATATATATATAAATATATAT-ATATGAAAATATAT 4413

Qy       503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
             ||||| | ||||||| |   | ||||| | ||| ||| | |||||||  | | |||||||
Db      4414 ATATAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATAT 4473
```

```
Qy        563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
              | ||||||| |||| | | ||||||||| |||||| | ||  | | ||||||| | ||
Db       4474 AAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAAT 4533

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
              ||||| ||||  | | | ||||||| | ||| |||||   | | ||| ||||| ||||||
Db       4534 ATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAATATAT 4593

Qy        683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742
              |||   | |||||||  || |||||| |||  | ||||       | |  |||      |
Db       4594 ATATGAAAATATATATATAAATATATATATGAAAATATATATATAAATATATATATATGA 4653

Qy        743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802
                |  |   ||  | |    ||| | ||||| | |  |  | |   | |     | ||
Db       4654 AAATATATATATAAATATATATATATGAAAATATATATAAATATATATATGAAAATATAT 4713

Qy        803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862
              | || |     |||    | |||     | ||||| ||| ||   ||   | |   |
Db       4714 ATATATATATATATGAAAATATATATATATGAAAATATATATATAAATATATATATGAAA 4773

Qy        863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATA 902
              |   |  ||   |  |  |||   |||||| | |||||
Db       4774 ATAATATATATAAATATATATAAATATATATAAAAATATA 4813
```

```
RESULT 14
ABK69933
ID   ABK69933 standard; DNA; 20420 BP.
XX
AC   ABK69933;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human secreted protein gene 68 genomic DNA fragment #21.
XX
KW   Human; ds; secreted protein; gene therapy; immunosuppressive;
KW   antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW   vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW   virucide; fungicide; opthalmalogical; autoimmune disease; neoplasm;
KW   rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW   cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW   angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW   ocular disorder; corneal infection; wound healing; skin aging;
KW   epithelial cell proliferation; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200226931-A2.
XX
PD   04-APR-2002.
XX
PF   24-SEP-2001; 2001WO-US29871.
XX
PR   25-SEP-2000; 2000US-234925P.
PR   12-JAN-2001; 2001WO-US00911.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis G,  Duan DR,  Rosen CA,  Moore PA,  Shi Y;
PI   Lafleur DW,  Olsen H,  Brewer LA,  Florence KA,  Young PE,  Soppet DR;
PI   Endress GA,  Mucenski M,  Ebner R;
XX
DR   WPI; 2002-362489/39.
XX
PT   Novel 71 isolated secreted polypeptides and polynucleotides encoding
PT   the polypeptides, useful for treating Huntington's disease, sepsis,
PT   meningitis, thrombocytopaenia, haemolytic anaemia, rheumatoid arthritis,
PT   asthma  -
XX
PS   Example 2; Page 1440-1445; 1478pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (or its
CC   fragment, homologue complement or allelic variant) encoding a human
CC   secreted protein (and its fragment, domain, epitope, variant, secreted
CC   form and species variant). Also included are a recombinant vector
CC   comprising the nucleic acid, a recombinant host cell comprising the
CC   vector, an antibody against the secreted protein, a recombinant host cell
CC   that expresses the secreted protein and a method of identifying a binding
CC   partner of the secreted protein. The nucleic acid and protein are used to
CC   prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC   humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep
CC   for example autoimmune diseases e.g. rheumatoid arthritis,
CC   hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC   cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC   e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC   Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC   ocular disorders e.g. corneal infection. Many other diseases and
CC   disorders are listed in the specification. The polypeptides can also be
CC   used to aid wound healing an epithelial cell proliferation, to prevent
CC   skin aging due to sunburn, to maintain organs before transplantation, for
CC   supporting cell culture of primary tissues, to regenerate tissues and in
CC   chemotaxis. The polypeptides can also be used as a food additive or
CC   preservative to increase or decrease storage capabilities. The present
CC   sequence represents a ds DNA fragment of the gene for a novel human
CC   secreted protein of the invention.
XX
SQ   Sequence 20420 BP; 4834 A; 5403 C; 5630 G; 4553 T; 0 other;
```

```
  Query Match             27.0%;  Score 270.4;  DB 24;  Length 20420;
  Best Local Similarity   61.4%;  Pred. No. 1.3e-34;
  Matches  467;  Conservative    0;  Mismatches  291;  Indels    2;  Gaps    2;

Qy        144 GGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTC 203
              | | | | || |  || ||  ||   |||| |||||      |         |
Db       4055 GCTGGAGTGCAGTAGCATGATCACAGCTCACTGCAGCCATATATATGTATATATATATAT 4114

Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              || | | |||| | | |      | |||||| | | |||||||||| | | | | |
Db       4115 ATATGAAAAAATATATATATATATATATATGAAAAAATATATATAAATATATATATGA 4174

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACAT-ATATACACATA 322
              | |||||||||||| | |||||||  | |||||| | ||||| | || ||||| | | |
Db       4175 AAATATATATATAAATATATATATGAAAATATATATAAATATATATATAATATATATAAA 4234

Qy        323 TATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATA 382
              |||||| | ||| | || || | ||  | ||| | |||  |||| | ||| ||| | |||
Db       4235 TATATATATATATAAATATATATATGAAAATATATATAAATATATATATAAATATATATA 4294

Qy        383 TATACACACATATATACACATATATACACACATATATACACATATATACACATATATACA 442
              |  | | | ||||||  | |||||||   | |||||||  | | |||||  | ||||  | |
Db       4295 TGAAAATATATATATGAAAATATATATGAAAATATATATATAAATATATATATATGAAAA 4354

Qy        443 CATATACACATATATACACATATATACATATATACACATATATATAATATACACACATAT 502
              |||||  | |  |||||  | ||  | | |||||||  | ||||||||  ||||  | | ||||
Db       4355 TATATATATAAATATATATATGAAAATATATATATAAATATATAT-ATATGAAAATATAT 4413

Qy        503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
              |||||  | |||||||  |   | |||||  | |||  |||  | ||||||   | |  |||||||
Db       4414 ATATAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATAT 4473

Qy        563 ACATATATACATATATACACATATATATACATATATACACATATATACACATATATACACAT 622
              | ||||||| ||||  | | |||||||| ||||||| | ||  | | ||||||| | ||
Db       4474 AAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAAT 4533

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
              ||||| ||||  | | | ||||||| | ||| |||||   | | ||| ||||| ||||||
Db       4534 ATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAATATAT 4593

Qy        683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742
              |||   | |||||||  || |||||| |||  | ||||       | |  |||      |
Db       4594 ATATGAAAATATATATATAAATATATATATGAAAATATATATATAAATATATATATATGA 4653

Qy        743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802
                |  |   ||  | |    ||| | ||||| | |  |  | |   | |     | ||
Db       4654 AAATATATATATAAATATATATATATGAAAATATATATAAATATATATATGAAAATATAT 4713
```

```
Qy        803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862
              | || |     |||    | |||     | ||||| ||| ||   ||   | |   |
Db       4714 ATATATATATATATGAAAATATATATATATGAAAATATATATATAAATATATATATGAAA 4773

Qy        863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATA 902
              |  |  ||   |   |  |||   |||||| | |||||
Db       4774 ATAATATATATAAATATATATAAATATATATAAAAATATA 4813
```

RESULT 15
AAK73166
ID AAK73166 standard; DNA; 19965 BP.
XX
AC AAK73166;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27978.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.

PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 27978; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 19965 BP; 4699 A; 5296 C; 5505 G; 4465 T; 0 other;

Query Match 26.7%; Score 267.4; DB 22; Length 19965;
Best Local Similarity 63.9%; Pred. No. 3.9e-34;
Matches 453; Conservative 0; Mismatches 251; Indels 5; Gaps 3;

```
Qy       215 AAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATATA 274
Db      4148 AAAAAATATATATAAATATATATATGAAAATATATATATAAATATATATATGAAAATATA 4207

Qy       275 TACACATATATATACACATATATACATATATACACATATATACACATATATATACACATA 334
Db      4208 TATAAATATATATATGAAAATATATATATGAAAATATATATAAATATATATATGAAAATA 4267

Qy       335 CATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATA 394
Db      4268 TATATGAAAATATATATATGAAAATATATATGAAAATATATATATAAATATATATATATG 4327

Qy       395 TATACACATATATACACACATATATACACATATATACACATATATACACATATACACATA 454
Db      4328 AAAATATATATATAAATATATATATGAAAATATATATATAAATATATATATGAAAATATA 4387

Qy       455 TATACACATATATACATATATACACATATATAT-AATATACACACATATATATACACATA 513
Db      4388 TATATAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATA 4447

Qy       514 TATACACACATATATACACATATATACACATATATATACACATATATACACATATATACA 573
Db      4448 TAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAA 4507

Qy       574 TATATACACATATATATACATATATACACATATATACATATATACACATATATACATATA 633
Db      4508 TATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAATATATA 4567

Qy       634 TACACACATATATACACATACATATACACACACATAGATATACATATATATACACATATA 693
Db      4568 TATATGAAAATATATATAAATATATATATGAATATATATATATATATATGAAAATATATA 4627

Qy       694 TATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTA 753
Db      4628 TATATGAAAATATATATATAAATATATATATGAAAATATATATAAATATATATATGAATA 4687

Qy       754 CCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAG 813
Db      4688 TATATATATATATATGAAAATATATATATATGAAAATATATATATAAATATATATATGAA 4747

Qy       814 GTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAG 873
Db      4748 AATAAT---ATATATAAATATATATAAATATATATAAAAATATAAATATATATAA-AAAT 4803

Qy       874 TTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTA 922
Db      4804 ATATAAATATATATATATAAACATATATATATAAATATATATATATATA 4852
```

Search completed: November 7, 2003, 06:15:31
Job time : 314.835 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 6, 2003, 21:13:51 ; Search time 621.98 Seconds
(without alignments)
8684.478 Million cell updates/sec

Title: US-09-939-209A-3_COPY_3000_5000
Perfect score: 2001
Sequence: 1 atttgtataattagcacctc.....cagttcttctaaagcctatt 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
✔	1	2001	100.0	20300	24	ABK47337	Genomic nucleotide
	2	173.8	8.7	126512	24	ABN83429	Human transporter
c	3	170	8.5	965	22	AAH33769	Human colon cancer
	4	167.4	8.4	168575	22	AAH21613	Human hypocretin r
c	5	164.2	8.2	541	25	ABZ71922	Human cancer-assoc
c	6	164.2	8.2	462586	25	ABQ84281	Chromosome 2 Asthm
	7	161.6	8.1	1503841	24	ABT00010	Human neuregulin 1
	8	161.6	8.1	1503841	24	ABT01503	Human neuregulin 1
	9	161.6	8.1	1503900	22	AAK95240	Human neuregulin-1
	10	161.6	8.1	1503900	22	AAK96733	Human neuregulin-1
c	11	161.2	8.1	154902	24	ABQ88198	Human osteoblast d
c	12	160	8.0	402	22	AAK62341	Human immune/haema
c	13	157.8	7.9	1332	21	AAC79685	Human secreted pro
	14	155.8	7.8	700	22	AAH92899	Human inflammatory
c	15	155.6	7.8	2604	20	AAX35699	cDNA encoding a pr
c	16	155.6	7.8	4167	22	AAK84449	Human immune/haema
	17	155.6	7.8	84607	20	AAX90847	Human PACAP genomi
c	18	154.8	7.7	680	22	AAS32407	Human cDNA encodin
c	19	154.6	7.7	139904	24	ABK83562	Human cDNA differe
c	20	154	7.7	22651	22	AAK78202	Human immune/haema
c	21	153	7.6	479	23	ABV56849	Human prostate exp
	22	151.6	7.6	45716	24	ABA93401	Human rTS-alpha ge
	23	151.6	7.6	45989	24	ABA93402	Human rTS-beta gen
	24	151.2	7.6	40352	19	AAV02032	MAGE-B cluster DNA
	25	151	7.5	6264	22	AAS32706	Human genomic DNA
	26	151	7.5	6266	22	AAS32708	Human genomic DNA
	27	151	7.5	6268	22	AAS32707	Human genomic DNA
c	28	149.4	7.5	35641	24	ABN95727	Gene #2225 used to
c	29	149.4	7.5	35641	24	ABL64428	Stomach cancer rel
c	30	149.4	7.5	44848	21	AAA75080	Nucleotide sequenc
	31	149	7.4	3271	22	ABA07229	Human pancreatic c
	32	149	7.4	3271	22	AAK89870	Human digestive sy
	33	148.4	7.4	106323	25	ABX14591	Human chloride int
c	34	148.2	7.4	4275	22	ABA19408	Human nervous syst
	35	148.2	7.4	84495	24	AAS20588	Human methionine a
c	36	148	7.4	1691080	24	ABX08336	Human phosphodiest
c	37	147.6	7.4	460	22	ABA07223	Human pancreatic c
c	38	147.6	7.4	460	22	AAK89864	Human digestive sy
c	39	147.6	7.4	1342	24	ABV72785	Phosphoribosyl gly
c	40	146.8	7.3	370	21	AAC23737	Human secreted pro
c	41	146.4	7.3	15439	22	AAD16515	Human ABC transpor
c	42	146.4	7.3	15439	22	AAS29744	Human endocrine po
	43	146.4	7.3	33513	22	AAK78746	Human immune/haema
	44	145.4	7.3	154902	24	ABQ88198	Human osteoblast d
c	45	145	7.2	172637	24	ABN83124	Human voltage-acti

ALIGNMENTS

RESULT 1
ABK47337
ID ABK47337 standard; DNA; 20300 BP.
XX
AC ABK47337;
XX
DT 18-JUN-2002 (first entry)
XX
DE Genomic nucleotide sequence encoding human RGS-4 protein.
XX
KW RGS-4; schizophrenia; human; regulator of G protein signalling 4;
KW neuroleptic; polymorphism; schizotypy; schizoid; chromosome 1q21-22;
KW gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace (4121,T)
FT /*tag= a
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4123,A)
FT /*tag= b
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4368,C)
FT /*tag= c
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4621,C)
FT /*tag= d
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4790,T)
FT /*tag= e

```
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (4816,T)
FT                   /*tag= f
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (4970,T)
FT                   /*tag= g
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace 5055,G)
FT                   /*tag= h
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (5295,A)
FT                   /*tag= i
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (5695,A)
FT                   /*tag= j
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (7375,T)
FT                   /*tag= k
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (7759,A)
FT                   /*tag= l
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (8596,A)
FT                   /*tag= m
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (9602..9610,CA)
FT                   /*tag= n
FT                   /note= "Deletion polymorphism"
FT   variation       replace (9892,A)
FT                   /*tag= o
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (9963,A)
FT                   /*tag= p
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (10132,A)
FT                   /*tag= q
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11056,C)
FT                   /*tag= r
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11091,T)
FT                   /*tag= s
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11106,A)
FT                   /*tag= t
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11774,T)
FT                   /*tag= u
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (12143,A)
FT                   /*tag= v
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (12145,T)
FT                   /*tag= w
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (14367,G)
FT                   /*tag= x
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (17027..17029,GA)
FT                   /*tag= y
FT                   /note= "Deletion polymorphism"
FT   variation       replace (17630,T)
FT                   /*tag= Z
FT                   /standard_name= "Single-nucleotide polymorphism"
XX
PN   WO200216653-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26622.
XX
PR   24-AUG-2000; 2000US-228021P.
XX
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Levitt PR, Mirnics K, Kodavali VC, Nimgaonkar VL;
XX
DR   WPI; 2002-292070/33.
XX
PT   Diagnosing, assessing susceptibility and treating schizophrenia,
PT   involves observing regulator of G-protein signalling 4, RGS4 levels in a
PT   subject -
XX
PS   Claim 1; Page 20-33; 112pp; English.
XX
CC   This invention relates to a novel method for diagnosing schizophrenia
CC   or determining susceptibility to schizophrenia in a human. The method
CC   comprises obtaining from a patient a DNA sample and detecting variations
CC   in the regulator of G-protein signalling 4 (RGS4) gene. Alternatively,
CC   the method involves measuring RGS4 mRNA or protein levels in a tissue
CC   sample from the patient and determining if there is a reduced level.
CC   The method of the invention is useful for diagnosing and determining
CC   susceptibility to schizophrenia. The invention also comprises a method
CC   that is useful for treating schizophrenia which includes a prophylactic
CC   treatment. The method of genotyping polymorphic variants in the RGS-4
CC   gene is applied to diagnosing pathologies of the schizophrenic spectrum,
CC   such as in particular schizotypy, schizoid individuals, etc. This
CC   method offers the possibility of diagnosing schizophrenia by a
CC   biological test and no longer exclusively by clinical evaluations.
CC   The present sequence represents the genomic DNA encoding the human
CC   regulator of G-protein signalling 4 (RGS4) protein used in the method of
CC   the invention. The gene for the RGS4 protein is located on human
CC   chromosome 1q21-22.
XX
SQ   Sequence 20300 BP; 6157 A; 4102 C; 3775 G; 6266 T; 0 other;

  Query Match           100.0%;  Score 2001;  DB 24;  Length 20300;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2001;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3000 ATTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAG 3059

Qy         61 AGAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3060 AGAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAG 3119

Qy        121 TCATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3120 TCATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGT 3179

Qy        181 CATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3180 CATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAA 3239

Qy        241 TATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3240 TATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGC 3299

Qy        301 AGGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3300 AGGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTA 3359

Qy        361 TAATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3360 TAATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAAT 3419

Qy        421 AGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3420 AGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTA 3479

Qy        481 CATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3480 CATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTC 3539
```

```
Qy       541 ACTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTG 600
Db      3540 ACTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTG 3599

Qy       601 CCCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTAC 660
Db      3600 CCCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTAC 3659

Qy       661 CCACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACC 720
Db      3660 CCACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACC 3719

Qy       721 TTTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCA 780
Db      3720 TTTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCA 3779

Qy       781 AATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACT 840
Db      3780 AATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACT 3839

Qy       841 TCTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTG 900
Db      3840 TCTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTG 3899

Qy       901 ACCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAA 960
Db      3900 ACCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAA 3959

Qy       961 TGTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAG 1020
Db      3960 TGTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAG 4019

Qy      1021 TGGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATG 1080
Db      4020 TGGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATG 4079

Qy      1081 CTAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAAC 1140
Db      4080 CTAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAAC 4139

Qy      1141 TAGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAAT 1200
Db      4140 TAGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAAT 4199

Qy      1201 CCAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCA 1260
Db      4200 CCAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCA 4259

Qy      1261 CCTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGC 1320
Db      4260 CCTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGC 4319

Qy      1321 AGAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAA 1380
Db      4320 AGAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAA 4379

Qy      1381 ACAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGA 1440
Db      4380 ACAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGA 4439

Qy      1441 TGCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCA 1500
Db      4440 TGCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCA 4499

Qy      1501 CAAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCC 1560
Db      4500 CAAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCC 4559

Qy      1561 TGTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAA 1620
Db      4560 TGTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAA 4619

Qy      1621 AATCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGG 1680
Db      4620 AATCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGG 4679

Qy      1681 GAGGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGC 1740
Db      4680 GAGGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGC 4739

Qy      1741 TGAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAG 1800
Db      4740 TGAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAG 4799

Qy      1801 AACTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATC 1860
Db      4800 AACTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATC 4859

Qy      1861 CTTGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACC 1920
Db      4860 CTTGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACC 4919

Qy      1921 TAGAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCCCAGTAGGAA 1980
Db      4920 TAGAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCCCAGTAGGAA 4979

Qy      1981 ACAGTTCTTCTAAAGCCTATT 2001
Db      4980 ACAGTTCTTCTAAAGCCTATT 5000
```

RESULT 2
ABN83429
ID ABN83429 standard; DNA; 126512 BP.
XX
AC ABN83429;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human transporter protein gene.
XX
KW Human; sodium/calcium exchanger; transporter; brain; heart; kidney; lung;
KW spleen; testis; leukocyte; foetal brain; chromosome 14; gene;
KW single nucleotide polymorphism; SNP; ds.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	variation	replace(378,T)
FT		/*tag= a
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(741..742,C-)
FT		/*tag= b
FT		/standard_name= "Single nucleotide polymorphism"
FT		/note= "This variation is an indel"
FT	variation	replace(2002,T)
FT		/*tag= c
FT		/standard_name= "Single nucleotide polymorphism"
FT	CDS	2010..124505
FT		/*tag= d
FT		/product= "Human transporter"
FT		/note= "Contains 5 introns"
FT	exon	2010..3793
FT		/*tag= e
FT		/number= 1
FT	variation	replace(2381,C)
FT		/*tag= f
FT		/standard_name= "Single nucleotide polymorphism"
FT	intron	3794..109509
FT		/*tag= g
FT		/number= 1
FT	variation	replace(5165,T)
FT		/*tag= h
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(5402,G)
FT		/*tag= i

```
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6794,C)
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9883,G)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10210,C)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(12220,G)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13842,G)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(14200,A)
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(15878,T)
FT                   /*tag= p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(16030,G)
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(16292,C)
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(16506,G)
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17953,A)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(23832,G)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25001,A)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25141,G)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25191,G)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(26147,A)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27400,G)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27401,T)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29278,C)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31437,G)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31857,G)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(33155,A)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(39487,C)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(41449,C)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(42420,C)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(43256,C)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(43967,C)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(48603..48604,A-)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "This variation is an indel"
FT   variation       replace(49560,T)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(52729,G)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(55031,G)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(55066,C)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(56912,G)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(58480,T)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(61128,A)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(61320,A)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(61444,C)
FT                   /*tag= at
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(62641,C)
FT                   /*tag= au
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(63023,G)
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(63051,C)
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(64989,G)
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(65929,A)
FT                   /*tag= ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(66694,G)
FT                   /*tag= az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(66755,A)
FT                   /*tag= ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(66879,C)
FT                   /*tag= bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(69156,T)
FT                   /*tag= bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(69280,T)
FT                   /*tag= bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(70647,T)
FT                   /*tag= be
FT                   /standard_name= "Single nucleotide polymorphism"
```

```
FT   variation       replace(71867,T)
FT                   /*tag=  bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(71900,T)
FT                   /*tag=  bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(71901,A)
FT                   /*tag=  bh
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(72369,T)
FT                   /*tag=  bi
```

```
  Query Match             8.7%;  Score 173.8;  DB 24;  Length 126512;
  Best Local Similarity   83.4%;  Pred. No. 3.5e-34;
  Matches  221;  Conservative    0;  Mismatches   42;  Indels    2;  Gaps    2;

Qy        308 TACCTTTTCTATGTTTAGATATG-TTAGATACACAAATATATTTCATTGTGTTATAATTT 366
              | ||||||||| || |||||||| |||||||||| ||| | |||||||||  ||||
Db      57055 TGCCTTTTCTAGGTCTAGATATGTTTAGATACACACATACTTACCATTGTGTTCCAATTG 57114

Qy        367 CCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAGGCTA 426
              |||||||| | |||||||||||| || ||||||||||||||||||||||| |||||||||
Db      57115 CCTACAGTTTCCAGTACAGTAACCTGTTGTACAGGTTTGTAACCTAGGAGCAATAGGCTA 57174

Qy        427 TACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACATTCT 486
              |||||||||||| |||||||||||||||||||| |||| ||||| || ||||||||| |||
Db      57175 TACCATACAGCCTAGGTGTGTAGTAGGCTAT-ACCACTTAGGTCTGGGTAAGTACACTCT 57233

Qy        487 ATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCACTGTT 546
              ||||| ||  |||| |||||||||    |||||  ||| ||||||||||||||| || | |||
Db      57234 ATGATGTTTTCACAGTGATGAAACTTCCTAATGACAAATTTCTCAGAATGTATCCCAGTT 57293

Qy        547 GTGAAGTGACCCATGACTATATTTT 571
              || ||||||  ||||||  || | |
Db      57294 GTTAAGTGAGGCATGACAGTACTAT 57318
```

```
RESULT 3
AAH33769/c
ID   AAH33769 standard; cDNA; 965 BP.
XX
AC   AAH33769;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen encoding cDNA SEQ ID NO:825.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999;   99US-0157137.
PR   03-NOV-1999;   99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   P-PSDB; AAG74338.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 1; Page 2780-2781; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAB77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence 965 BP; 315 A; 126 C; 174 G; 349 T; 1 other;
```

```
  Query Match             8.5%;  Score 170;  DB 22;  Length 965;
  Best Local Similarity   72.2%;  Pred. No. 5e-34;
  Matches  283;  Conservative    0;  Mismatches   90;  Indels   19;  Gaps    4;

Qy        182 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 241
              ||  | || | |   |||||||  |||   || || |||| |||      |  || ||||
Db        631 ATACGACATACAGTCATGTTTCGATCAACAATTGACCACATATGACAGAGATCCTATAAG 572

Qy        242 ATTATA----AGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAG--TGT 295
              ||||||    | |||| ||||| |||| | ||||||||  | |||||||  | |   |||
Db        571 ATTATAATGGAACTGAAAAATTCCTATCACCTAGTGATGCCACAGCCATTGTTACATTGT 512

Qy        296 AATGCAGGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTG 355
              |   ||    |||| || ||||||||||||||     ||||||            |||||
Db        511 AGCACAATGTATTATCTGTTCTATGTTTAGATTACACAGATAC-----------CATTG 464

Qy        356 TGTTATAATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGA 415
              ||||| | || |||||||||||||| ||||||| |||||||||||||| |||| || || |
Db        463 TGTTACAGTTGCCTACAGTATTCAATACAGTACCATGCTGTACAGGTCTGTAGCCCAGAA 404

Qy        416 GTAATAGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGT 475
              | | || || | ||||  |||  ||||||| |||||||| | |||| ||||||||||||
Db        403 GCGACAGACTTTGCCATCTAGCCCAGGTGTGCAGTAGGCTCT-ACCACCTAGGTTTGTGT 345

Qy        476 AAGTACATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAAT 535
              |||||||  |  ||||| ||| |||||||||||||||||||  ||| ||||||||||||
Db        344 AAGTACACTTCATGATGTTCACACAATGATGAAATCACCTAATGACAAATTTCTCAGAAT 285

Qy        536 GTTTCACTGTTGTGAAGTGACCCATGACTATA 567
              ||  | ||||  |||||| || ||| ||| ||
Db        284 GTAGCTCTGTCATGAAGTCACACATTACTGTA 253
```

```
RESULT 4
AAH21613
ID   AAH21613 standard; DNA; 168575 BP.
XX
AC   AAH21613;
XX
DT   10-AUG-2001  (first entry)
XX
DE   Human hypocretin receptor 2 (HCRTR2) gene SEQ ID NO:1.
XX
KW   Human; narcolepsy; hypocretin receptor 2; orexin receptor 2; HCRTR2;
KW   diagnosis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200130991-A2.
XX
PD   03-MAY-2001.
```

```
XX
PF   22-AUG-2000; 2000WO-US23021.
XX
PR   25-OCT-1999;   99US-0426290.
XX
PA   (DECO-) DECODE GENETICS EHF.
XX
PI   Olafsdottir BR,  Gulcher J;
XX
DR   WPI; 2001-300504/31.
DR   P-PSDB; AAB98007.
XX
PT   Gene for hypocretin (orexin) receptor 2 (HCRTR2) which is associated
PT   with narcolepsy, useful in methods of diagnosis of narcolepsy and
PT   pharmaceutical compositions for therapy -
XX
PS   Claim 1; Fig 1; 85pp; English.
XX
CC   The present invention describes the human hypocretin (orexin) receptor 2
CC   (HCRTR2) gene (given in AAH21613), which is associated with narcolepsy.
CC   Identification of the HCRTR2 nucleic acid molecule permits the diagnosis
CC   of narcolepsy. A method from the present invention is provided for
CC   treating narcolepsy by administering to the individual an isolated
CC   HCRTR2 nucleic acid in a therapeutically effective amount so that the
CC   cells produce native HCRTR2 receptor. The diagnosis of narcolepsy has
CC   been difficult to differentiate from other conditions such as chronic
CC   fatigue syndrome or other sleep disorders but detection of HCRTR2
CC   nucleic acid makes it possible to accurately diagnose narcolepsy.
CC   AAH21541 to AAH21612 represent primers used in the identification of the
CC   narcolepsy gene in an example from the present invention. AAH21613
CC   represents the HCRTR2 gene which encodes the HCRTR2 protein given in
CC   AAB98007.
XX
SQ   Sequence 168575 BP; 55308 A; 29672 C; 29838 G; 53757 T; 0 other;

  Query Match             8.4%;  Score 167.4;  DB 22;  Length 168575;
  Best Local Similarity   79.2%;  Pred. No. 1.8e-32;
  Matches  224;  Conservative    0;  Mismatches   56;  Indels    3;  Gaps    2;

Qy        311 CTTTTCTATGTTTAGATATGTT-AGATACACAAATATATTTCATTGTGTTATAATTTCCT 369
              |||||||||||||||||||||| ||||||||||||   | |||||||||||||||| |||
Db      57581 CTTTTCTATGTTTAGATATGTTCAGATACACAAATGCTTATCATTGTGTTATAATTGCCT 57640

Qy        370 ACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAGGCTATAC 429
              |||||  |||||||||| |||||||||||||||| || |||||||| ||| ||||||||
Db      57641 ACAGTGTTCAGTACAGTTACATGCTGTACAGGTTTATAGCCTAGGAGCAATTGGCTATAC 57700

Qy        430 CATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACATTCTATG 489
              | || ||| ||||||||||||||||||||   |   ||| |||||||||| | |  |||||
Db      57701 CCTATAGCCTAGGTGTGTAGTAGGCTATA--CCATTAGATTTGTGTAAGCATACCCTATG 57758

Qy        490 ATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCACTGTTGTG 549
              || ||  |||||||||||||||||||||  |  |||||||||| || | || | ||  |
Db      57759 ATGTTTGCACAATGATGAAATCACCTAAGGATGCATTTCTCAGCATATATCCCAGTCATT 57818

Qy        550 AAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTG 592
              |||  |    ||||| |||| |    | || || ||  | | |
Db      57819 AAGCAAAGACTGACTCTATTATTAGGTCTATTTTATTCTATAG 57861


RESULT 5
ABZ71922/c
ID   ABZ71922 standard; cDNA; 541 BP.
XX
AC   ABZ71922;
XX
DT   01-APR-2003  (first entry)
XX
DE   Human cancer-associated gene SEQ ID NO 229.
XX
KW   Human; cancer; stomach cancer; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200283899-A1.
XX
PD   24-OCT-2002.
XX
PF   28-MAR-2002; 2002WO-JP03038.
XX
PR   10-APR-2001; 2001JP-0112039.
PR   21-SEP-2001; 2001JP-0290193.
XX
PA   (TAKA-) TAKARA BIO INC.
XX
PI   Yoshikawa Y,  Okamoto S,  Oura T,  Mineno J,  Asada K,  Kato I;
PI   Inoue H,  Mori M;
XX
DR   WPI; 2003-093022/08.
XX
PT   Measuring changes in expression of 264 cancer associated genes for
PT   detection of stomach cancer and screening of potential anticancer
PT   agents -
XX
PS   Claim 1; Page 235-236; 266pp; Japanese.
XX
CC   The invention relates to a method for the detection of cancer in which a
CC   change in the expression of 1 or more of 264 specified cancer associated
CC   genes, ABZ71694-ABZ71957, or of sequences at least 80% homologous to them
CC   in the specimen tissue as compared to normal tissue is observed. The
CC   genes are used in detection, diagnosis and treatment of cancer,
CC   especially of stomach cancer. The present sequence is that of a cancer
CC   associated polynucleotide of the invention.
XX
SQ   Sequence 541 BP; 166 A; 110 C; 117 G; 148 T; 0 other;

  Query Match             8.2%;  Score 164.2;  DB 25;  Length 541;
  Best Local Similarity   81.1%;  Pred. No. 1.3e-32;
  Matches  215;  Conservative    0;  Mismatches   48;  Indels    2;  Gaps    2;

Qy        304 ACATTACCTTTTCTATGTTTAGATAT-GTTAGATACACAAATATATTTCATTGTGTTATA 362
              || |||||||||||||||||||| | | | |||||| |||||||  |  ||||||||| ||
Db        395 ACCTTACCTTTTCTATGTTTAGCTCTGGGTAGATAGACAAATACTTACCATTGTGTTTTA 336

Qy        363 ATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAG 422
              ||| ||| || |||||||||||||||||||||||||| | ||| |||||||  || ||
Db        335 ATTGCCTGCAATATTCAGTACAGTAACATGCTGTACAGGCTCGTAGCCTAGGACCAACAG 276

Qy        423 GCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACA 482
              |||||||||| ||| ||||| |||||||||||| |||||||||||||||||||||||||
Db        275 GCTATACCATATAGCCTAGGTATGTAGTAGGCTAT-ACCATCTAGGTTTGTGTAAGTACA 217

Qy        483 TTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCAC 542
              ||| ||||| |  |||||   ||||||  || ||  |||||||||||||| |  | |  |
Db        216 CTCTTTGATACTTGCACAACCTTGAAATTGCCGAATGACACATTTCTCAGGACATATTCC 157

Qy        543 TGTTGTGAAGTGACCCATGACTATA 567
              |||    |||  |||||||||| ||
Db        156 TGTCACCAAGCAACCCATGACTGTA 132


RESULT 6
ABQ84281/c
ID   ABQ84281 standard; DNA; 462586 BP.
XX
AC   ABQ84281;
XX
DT   20-FEB-2003  (first entry)
XX
DE   Chromosome 2 Asthma Locus DNA sequence.
XX
KW   DPP10; dipeptidyl peptidase; prolyloligopeptidase; enzyme; asthma;
KW   antiinflammatory; antiasthmatic; antipsoriatic; antiarthritic;
KW   antirheumatic; vaccine; gene therapy; inflammatory disease;
KW   inflammatory bowel disease; atopy; rheumatoid arthritis; psoriasis;
KW   chromosome 2q14; gene; ds.
```

```
XX
OS   Homo sapiens.
XX
PN   WO200286113-A2.
XX
PD   31-OCT-2002.
XX
PF   24-APR-2002; 2002WO-GB01887.
XX
PR   24-APR-2001; 2001GB-0010044.
PR   24-APR-2001; 2001GB-0010046.
PR   12-OCT-2001; 2001GB-0024575.
PR   12-OCT-2001; 2001GB-0024594.
XX
PA   (ISIS-) ISIS INNOVATIONS LTD.
XX
PI   Cookson WOCM,  Moffat MF,  Allen M,  Lench N;
XX
DR   WPI; 2003-093132/08.
XX
PT   New nucleic acid sequence comprising DPP10 mRNA, useful for the
PT   manufacture of a medicament for regulating DPP10 protein expression or
PT   for preventing or treating inflammatory disease e.g., inflammatory
PT   bowel disease -
XX
PS   Claim 39; Fig 1; 321pp; English.
XX
CC   The present invention describes a new isolated nucleic acid sequence (I)
CC   comprising a DPP10 mRNA sequence. DPP10 is a dipeptidyl peptidase (also
CC   known as prolyloligopeptidase). (I) has antiinflammatory, antiasthmatic,
CC   antipsoriatic, antiarthritic and antirheumatic activities, and can be
CC   used in vaccines and gene therapy. A composition comprising (I) can be
CC   used for the manufacture of a medicament for regulating DPP10 expression
CC   or for preventing or treating inflammatory disease e.g., inflammatory
CC   bowel disease, asthma, atopy, rheumatoid arthritis or psoriasis. (I) can
CC   also be used in an assay for detecting or measuring DPP10 in a sample.
CC   A host cell comprising (I) can be used for producing recombinant DPP10
CC   gene products, or in drug screening systems to identify agents for
CC   diagnosis or treatment of individuals having or susceptible to
CC   inflammatory disease. Human DPP10 is located on chromosome 2, more
CC   specifically chromosome 2q14. ABQ84254 to ABQ84612 and ABP55569 to
CC   ABP55629 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 462586 BP; 136666 A; 87684 C; 89317 G; 148885 T; 34 other;

  Query Match             8.2%;  Score 164.2;  DB 25;  Length 462586;
  Best Local Similarity   63.1%;  Pred. No. 1.8e-31;
  Matches  286;  Conservative    0;  Mismatches  163;  Indels    4;  Gaps    2;

Qy        124 TCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTCAT 183
              ||||| ||  | |  |||   | | | ||| || |||    | | |    |    |
Db     333975 TCCGTCTTCAAGTAAGTTCCTGATTTAATCCCACTTTGCATTGACACTCCACTCTCCATT 333916

Qy        184 GTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAATAT 243
                |  |   |   || | |  | || | ||||   |||| |||        | |
Db     333915 TATGAAGGTTCTATCTCTTTATTGTGTGAGACAACAGTCATGTGTCTCAGAAAGACGTTT 333856

Qy        244 TATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCAGG 303
              |  |       |||       ||    |  |   |     || ||||    |     |
Db     333855 TGGACAAAATCAAACCATATATATGACGGTAGTCCCATAAGATTATAACACAATGTTTTG 333796

Qy        304 ACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTATAA 363
              ||  |  |||||||||||| |||||||||    || | |  |  | |||||||| || ||
Db     333795 ACTATTTCTTTTCTATGTTCAGATATGTT---TATATACGTGCTTATCATTGTGCTACAA 333739

Qy        364 TTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAGG 423
              || ||||||||||||||||||||||||||||||||||||||| | |||||| ||||||
Db     333738 TTGCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTATCTTAGGAGCAATAGG 333679

Qy        424 CTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACAT 483
              || ||||||| | | ||||||||||||| |||||| |||||||||||||||||| ||| |
Db     333678 CTGTACCATATAACCTAGGTGTGTAGTGGGCTAT-ACCATCTAGGTTTGTGTATGTATAC 333620

Qy        484 TCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCACT 543
              ||| |||| ||  || |||||  ||||| |||||| |||||||||||||||  | || ||
Db     333619 TCTGTGATGTTTGCATAATGACAAAATCGCCTAACAACACATTTCTCAGAACATATCTCT 333560

Qy        544 GTTGTGAAGTGACCCATGACTATATTTTCCTAT 576
              ||  | ||  |   ||| |||||| |||   ||
Db     333559 GTCATTAAACGGTGCATAACTATACTTTGTAAT 333527


RESULT 7
ABT00010
ID   ABT00010 standard; DNA; 1503841 BP.
XX
AC   ABT00010;
XX
DT   07-NOV-2002  (first entry)
XX
DE   Human neuregulin 1 gene.
XX
KW   Human; neuregulin 1; neuregulin-1-associated gene 1; NRG1; NRG1AG1;
KW   schizophrenia; chromosome 8p12; single nucleotide polymorphism; SNP;
KW   neuroleptic; gene therapy; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             244312..1369465
FT                   /*tag= a
FT                   /product= "neuregulin 1"
FT                   /note= "this sequence contains introns"
FT   exon            244205..244348
FT                   /*tag= b
FT                   /number= 1
FT   intron          244349..244640
FT                   /*tag= c
FT                   /number= 1
FT   exon            244641..245646
FT                   /*tag= d
FT                   /number= 2
FT   intron          245647..826009
FT                   /*tag= e
FT                   /number= 2
FT   exon            826010..826101
FT                   /*tag= f
FT                   /number= 3
FT   intron          826102..826307
FT                   /*tag= g
FT                   /number= 3
FT   exon            826308..826355
FT                   /*tag= h
FT                   /number= 4
FT   intron          826356..1034242
FT                   /*tag= i
FT                   /number= 4
FT   exon            1034243..1034321
FT                   /*tag= j
FT                   /number= 5
FT   intron          1034322..1153294
FT                   /*tag= k
FT                   /number= 5
FT   exon            1153295..1153886
FT                   /*tag= l
FT                   /number= 6
FT   intron          1153887..1200721
FT                   /*tag= m
FT                   /number= 6
FT   exon            1200722..1201065
FT                   /*tag= n
FT                   /number= 7a
FT   intron          1153887..1200887
FT                   /*tag= o
FT                   /number= 7a
```

```
FT   exon            1200888..1201065
FT                   /*tag= p
FT                   /number= 7b
FT   intron          1201066..1210622
FT                   /*tag= q
FT                   /number= 7b
FT   exon            1210623..1210744
FT                   /*tag= r
FT                   /number= 8
FT   intron          1210745..1219542
FT                   /*tag= s
FT                   /number= 8
FT   exon            1219543..1219593
FT                   /*tag= t
FT                   /number= 9
FT   intron          1219594..1221863
FT                   /*tag= u
FT                   /number= 9
FT   exon            1221864..1221914
FT                   /*tag= v
FT                   /number= 10
FT   intron          1221915..1252253
FT                   /*tag= w
FT                   /number= 10
FT   exon            1252254..1253413
FT                   /*tag= x
FT                   /number= 11
FT   intron          1253414..1326781
FT                   /*tag= y
FT                   /number= 11
FT   exon            1326782..1327071
FT                   /*tag= z
FT                   /number= 12
FT   intron          1327072..1332977
FT                   /*tag= aa
FT                   /number= 12
FT   exon            1332978..1333107
FT                   /*tag= ab
FT                   /number= 13a
FT   intron          1333108..1347039
FT                   /*tag= ac
FT                   /number= 13a
FT   exon            1332978..1333652
FT                   /*tag= ad
FT                   /number= 13b
FT   intron          1333653..1347039
FT                   /*tag= ae
FT                   /number= 13b
FT   exon            1347040..1347107
FT                   /*tag= af
FT                   /number= 14
FT   intron          1347108..1347706
FT                   /*tag= ag
FT                   /number= 14
FT   exon            1347707..1347765
FT                   /*tag= ah
FT                   /number= 15a
FT   intron          1347766..1354620
FT                   /*tag= aj
FT                   /number= 15a
FT   exon            1347707..1348257
FT                   /*tag= ak
FT                   /number= 15b
FT   intron          1348258..1354620
FT                   /*tag= al
FT                   /number= 15b
FT   exon            1354621..1354644
FT                   /*tag= am
FT                   /number= 16
FT   intron          1354645..1359431
FT                   /*tag= an
FT                   /number= 16
FT   exon            1359432..1359534
FT                   /*tag= ao
FT                   /number= 17
FT   intron          1359535..1361462
FT                   /*tag= ap
FT                   /number= 17
FT   exon            1361463..1361589
FT                   /*tag= aq
FT                   /number= 18
FT   intron          1361590..1364362
FT                   /*tag= ar
FT                   /number= 18
FT   exon            1364363..1364493
FT                   /*tag= as
FT                   /number= 19
FT   intron          1364494..1365256
FT                   /*tag= at
FT                   /number= 19
FT   exon            1365257..1365463
FT                   /*tag= au
FT                   /number= 20a
FT   intron          1365464..1368280
FT                   /*tag= av
FT                   /number= 20a
FT   exon            1365257..1366044
FT                   /*tag= aw
FT                   /number= 20b
FT   intron          1366045..1368280
FT                   /*tag= ax
FT                   /number= 20b
FT   exon            1368281..1368422
FT                   /*tag= ay
FT                   /number= 21
FT   intron          1368423..1368810
FT                   /*tag= az
FT                   /number= 21
FT   exon            1368811..1369656
FT                   /*tag= ba
FT                   /number= 22
FT   CDS             629728..744437
FT                   /*tag= bb
FT                   /product= "neuregulin-1-associated gene 1"
FT                   /note= "this sequence contains introns"
FT   exon            629728..629804
FT                   /*tag= bc
FT                   /number= 1
FT   intron          629805..630974
FT                   /*tag= bd
FT                   /number= 1
FT   exon            630975..631187
FT                   /*tag= be
FT                   /number= 2
FT   intron          631188..631282
FT                   /*tag= bf
FT                   /number= 2
FT   exon            631283..631320
FT                   /*tag= bg
FT                   /number= 3
FT   intron          631321..634334
FT                   /*tag= bh
FT                   /number= 3
FT   exon            634335..634441
FT                   /*tag= bi
FT                   /number= 4a
FT   intron          634442..635331
FT                   /*tag= bj
FT                   /number= 4a
```

Query Match 8.1%; Score 161.6; DB 24; Length 1503841;
Best Local Similarity 82.5%; Pred. No. 1.4e-30;
Matches 221; Conservative 0; Mismatches 44; Indels 3; Gaps 3;

```
Qy      304 ACATTACCTTTTCTATGTTTAGATATG-TTAGATACACAAATATATTTCATTGTGTTATA 362
            ||  |||||||||||||||||| |||| |||||| ||||||||| |  |||| |||| ||
```

```
Db      73826 ACTGTACCTTTTCTATGTTTAGTTATGTTTAGATTCACAAATATTTCCCATTTTGTTGTA 73885

Qy        363 ATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAG 422
              ||| ||||||||||||||||||||||||||||||  ||||||| |||||||| |||||
Db      73886 ATTGCCTACAGTATTCAGTACAGTAACATGCTGTACCAGTTTGTAGCCTAGGAGCAATAG 73945

Qy        423 GCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACA 482
              || ||||||| ||| ||  | || | |||||||  || ||||||||||||||||||| ||
Db      73946 ACTCTACCATATAGCCTACATCTGCAATAGGCTA-CACTATCTAGGTTTGTGTAAGTTCA 74004

Qy        483 TTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCAC 542
              |||||||| ||| | |||||| |||||||||  |  |||||||||||||| || || |
Db      74005 CTCTATGATGTTCACCCAATGAAGAAATCACCCGA-GACACATTTCTCAGAACGTATCCC 74063

Qy        543 TGTTGTGAAGTGACCCATGACTATATTT 570
              |||  | ||||||  ||||||| || ||
Db      74064 TGTCATTAAGTGAAACATGACTGTAGTT 74091
```

RESULT 8
ABT01503
ID ABT01503 standard; DNA; 1503841 BP.
XX
AC ABT01503;
XX
DT 07-NOV-2002 (first entry)
XX
DE Human neuregulin 1 gene.
XX
KW Human; neuregulin 1; neuregulin-1-associated gene 1; NRG1; NRG1AG1;
KW schizophrenia; chromosome 8p12; single nucleotide polymorphism; SNP;
KW neuroleptic; gene therapy; gene; ds.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	*CDS*	244312..1369465
FT		/*tag= a
FT		/product= "neuregulin 1"
FT		/note= "this sequence contains introns"
FT	exon	244205..244348
FT		/*tag= b
FT		/number= 1
FT	intron	244349..244640
FT		/*tag= c
FT		/number= 1
FT	exon	244641..245646
FT		/*tag= d
FT		/number= 2
FT	intron	245647..826009
FT		/*tag= e
FT		/number= 2
FT	exon	826010..826101
FT		/*tag= f
FT		/number= 3
FT	intron	826102..826307
FT		/*tag= g
FT		/number= 3
FT	exon	826308..826355
FT		/*tag= h
FT		/number= 4
FT	intron	826356..1034242
FT		/*tag= i
FT		/number= 4
FT	exon	1034243..1034321
FT		/*tag= j
FT		/number= 5
FT	intron	1034322..1153294
FT		/*tag= k
FT		/number= 5
FT	exon	1153295..1153886
FT		/*tag= l
FT		/number= 6
FT	intron	1153887..1200721
FT		/*tag= m
FT		/number= 6
FT	exon	1200722..1201065
FT		/*tag= n
FT		/number= 7a
FT	intron	1153887..1200887
FT		/*tag= o
FT		/number= 7a
FT	exon	1200888..1201065
FT		/*tag= p
FT		/number= 7b
FT	intron	1201066..1210622
FT		/*tag= q
FT		/number= 7b
FT	exon	1210623..1210744
FT		/*tag= r
FT		/number= 8
FT	intron	1210745..1219542
FT		/*tag= s
FT		/number= 8
FT	exon	1219543..1219593
FT		/*tag= t
FT		/number= 9
FT	intron	1219594..1221863
FT		/*tag= u
FT		/number= 9
FT	exon	1221864..1221914
FT		/*tag= v
FT		/number= 10
FT	intron	1221915..1252253
FT		/*tag= w
FT		/number= 10
FT	exon	1252254..1253413
FT		/*tag= x
FT		/number= 11
FT	intron	1253414..1326781
FT		/*tag= y
FT		/number= 11
FT	exon	1326782..1327071
FT		/*tag= z
FT		/number= 12
FT	intron	1327072..1332977
FT		/*tag= aa
FT		/number= 12
FT	exon	1332978..1333107
FT		/*tag= ab
FT		/number= 13a
FT	intron	1333108..1347039
FT		/*tag= ac
FT		/number= 13a
FT	exon	1332978..1333652
FT		/*tag= ad
FT		/number= 13b
FT	intron	1333653..1347039
FT		/*tag= ae
FT		/number= 13b
FT	exon	1347040..1347107
FT		/*tag= af
FT		/number= 14
FT	intron	1347108..1347706
FT		/*tag= ag
FT		/number= 14
FT	exon	1347707..1347765
FT		/*tag= ah
FT		/number= 15a
FT	intron	1347766..1354620
FT		/*tag= aj
FT		/number= 15a
FT	exon	1347707..1348257
FT		/*tag= ak
FT		/number= 15b
FT	intron	1348258..1354620

```
FT                   /*tag=  al
FT                   /number= 15b
FT   exon            1354621..1354644
FT                   /*tag=  am
FT                   /number= 16
FT   intron          1354645..1359431
FT                   /*tag=  an
FT                   /number= 16
FT   exon            1359432..1359534
FT                   /*tag=  ao
FT                   /number= 17
FT   intron          1359535..1361462
FT                   /*tag=  ap
FT                   /number= 17
FT   exon            1361463..1361589
FT                   /*tag=  aq
FT                   /number= 18
FT   intron          1361590..1364362
FT                   /*tag=  ar
FT                   /number= 18
FT   exon            1364363..1364493
FT                   /*tag=  as
FT                   /number= 19
FT   intron          1364494..1365256
FT                   /*tag=  at
FT                   /number= 19
FT   exon            1365257..1365463
FT                   /*tag=  au
FT                   /number= 20a
FT   intron          1365464..1368280
FT                   /*tag=  av
FT                   /number= 20a
FT   exon            1365257..1366044
FT                   /*tag=  aw
FT                   /number= 20b
FT   intron          1366045..1368280
FT                   /*tag=  ax
FT                   /number= 20b
FT   exon            1368281..1368422
FT                   /*tag=  ay
FT                   /number= 21
FT   intron          1368423..1368810
FT                   /*tag=  az
FT                   /number= 21
FT   exon            1368811..1369656
FT                   /*tag=  ba
FT                   /number= 22
FT   CDS             629728..744437
FT                   /*tag=  bb
FT                   /product= "neuregulin-1-associated gene 1"
FT                   /note= "this sequence contains introns"
FT   exon            629728..629804
FT                   /*tag=  bc
FT                   /number= 1
FT   intron          629805..630974
FT                   /*tag=  bd
FT                   /number= 1
FT   exon            630975..631187
FT                   /*tag=  be
FT                   /number= 2
FT   intron          631188..631282
FT                   /*tag=  bf
FT                   /number= 2
FT   exon            631283..631320
FT                   /*tag=  bg
FT                   /number= 3
FT   intron          631321..634334
FT                   /*tag=  bh
FT                   /number= 3
FT   exon            634335..634441
FT                   /*tag=  bi
FT                   /number= 4a
FT   intron          634442..635331
FT                   /*tag=  bj
FT                   /number= 4a
```

```
  Query Match             8.1%;  Score 161.6;  DB 24;  Length 1503841;
  Best Local Similarity   82.5%;  Pred. No. 1.4e-30;
  Matches  221;  Conservative    0;  Mismatches   44;  Indels    3;  Gaps    3;

Qy        304 ACATTACCTTTTCTATGTTTAGATATG-TTAGATACACAAATATATTTCATTGTGTTATA 362
              ||  |||||||||||||||||| |||| |||||| ||||||||| |  |||| |||| ||
Db      73826 ACTGTACCTTTTCTATGTTTAGTTATGTTTAGATTCACAAATATTTCCCATTTTGTTGTA 73885

Qy        363 ATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAG 422
              ||| ||||||||||||||||||||||||||||||  ||||||| |||||||| |||||
Db      73886 ATTGCCTACAGTATTCAGTACAGTAACATGCTGTACCAGTTTGTAGCCTAGGAGCAATAG 73945

Qy        423 GCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACA 482
              || ||||||| ||| ||  | || | |||||||  || ||||||||||||||||| ||
Db      73946 ACTCTACCATATAGCCTACATCTGCAATAGGCTA-CACTATCTAGGTTTGTGTAAGTTCA 74004

Qy        483 TTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCAC 542
              |||||||| ||| | |||||| |||||||||  |  |||||||||||||| || || |
Db      74005 CTCTATGATGTTCACCCAATGAAGAAATCACCCGA-GACACATTTCTCAGAACGTATCCC 74063

Qy        543 TGTTGTGAAGTGACCCATGACTATATTT 570
              |||  | ||||||  ||||||| || ||
Db      74064 TGTCATTAAGTGAAACATGACTGTAGTT 74091
```

```
RESULT 9
AAK95240
ID   AAK95240 standard; DNA; 1503900 BP.
XX
AC   AAK95240;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human neuregulin-1 gene.
XX
KW   Human; neuregulin-1 associated gene 1; NRG1AG1; Schizophrenia gene;
KW   gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200164876-A2.
XX
PD   07-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US06376.
XX
PR   28-FEB-2000; 2000US-0515715.
XX
PA   (DECO-) DECODE GENETICS EHF.
XX
PI   Stefansson H,  Steinthorsdottir V,  Gulcher JR;
XX
DR   WPI; 2001-550179/61.
DR   P-PSDB; AAG67900, AAG67901, AAG67902, AAG67903, AAG67904, AAG67905,
DR   AAG67906, AAG67907, AAG67908, AAG67909, AAG67910, AAG67911, AAG67912,
DR   AAG67913, AAG67914, AAG67915, AAG67916, AAG67917, AAG67918, AAG67919,
DR   AAG67920, AAG67921, AAG67922, AAG67923, AAG67924, AAG67925, AAG67926,
DR   AAG67927, AAG67928, AAG67929, AAG67930, AAG67931, AAG67932, AAG67933,
DR   AAG67934, AAG67935, AAG67936, AAG67937.
XX
PT   Neuregulin-1 associated gene 1 nucleic acids and fragments, useful for
PT   preventing diagnosing and treating schizophrenia -
XX
PS   Disclosure; Page 90-501; 750pp; English.
XX
CC   This sequence represents the human neuregulin-1 associated gene 1
CC   (NRG1AG1) of the invention. The NRG1AG1 gene is also referred to as the
CC   human Schizophrenia gene. The invention also relates to fragments or
CC   variants of the gene and the NRG1AG1 polypeptides they encode. The
CC   NRG1AG1 nucleic acids and polypeptides may be used in the prevention,
```

```
CC   diagnosis and treatment of diseases associated with inappropriate NRG1AG1
CC   expression. For example, they may be used to treat disorders associated
CC   with decreased expression by rectifying mutations or deletions in a
CC   patient's genome that affect the activity of NRG1AG1 by expressing
CC   inactive proteins or to supplement the patients own production of
CC   NRG1AG1. Additionally, the gene may be used to produce NRG1AG1
CC   polypeptides, by inserting the nucleic acids into a host cell and
CC   culturing the cell to express the protein. The gene may also be used as
CC   DNA probes and primers in diagnostic assays to detect and quantitate the
CC   presence of similar nucleic acids in samples, and therefore which
CC   patients may be in need of restorative therapy. The NRG1AG1 polypeptides
CC   may also be used as antigens in the production of antibodies against
CC   NRG1AG1 and in assays to identify modulators of NRG1AG1 expression and
CC   activity. Anti-NRG1AG1 antibodies and antagonists may also be used to
CC   down regulate expression and activity. Anti-NRG1AG1 antibodies may
CC   also be used as diagnostic agents for detecting the presence of NRG1AG1
CC   polypeptides in samples. NRG1AG1 Is associated with schizophrenia which
CC   may be prevented, diagnosed and/or treated by the above methods.
XX
SQ   Sequence 1503900 BP; 452487 A; 281874 C; 288074 G; 480092 T; 1373 other;

  Query Match             8.1%;  Score 161.6;  DB 22;  Length 1503900;
  Best Local Similarity   82.5%;  Pred. No. 1.4e-30;
  Matches  221;  Conservative    0;  Mismatches   44;  Indels    3;  Gaps    3;

Qy        304 ACATTACCTTTTCTATGTTTAGATATG-TTAGATACACAAATATATTTCATTGTGTTATA 362
              ||  |||||||||||||||||| |||| |||||| ||||||||| |  |||| |||| ||
Db      73826 ACTGTACCTTTTCTATGTTTAGTTATGTTTAGATTCACAAATATTTCCCATTTTGTTGTA 73885

Qy        363 ATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAG 422
              ||| ||||||||||||||||||||||||||||||  ||||||| |||||||| |||||
Db      73886 ATTGCCTACAGTATTCAGTACAGTAACATGCTGTACCAGTTTGTAGCCTAGGAGCAATAG 73945

Qy        423 GCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACA 482
              || ||||||| ||| ||  | || | |||||||  || ||||||||||||||||||| ||
Db      73946 ACTCTACCATATAGCCTACATCTGCAATAGGCTA-CACTATCTAGGTTTGTGTAAGTTCA 74004

Qy        483 TTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCAC 542
              |||||||| ||| | |||||| |||||||||  |  |||||||||||||| || || |
Db      74005 CTCTATGATGTTCACCCAATGAAGAAATCACCCGA-GACACATTTCTCAGAACGTATCCC 74063

Qy        543 TGTTGTGAAGTGACCCATGACTATATTT 570
              |||  | ||||||  ||||||| || ||
Db      74064 TGTCATTAAGTGAAACATGACTGTAGTT 74091


RESULT 10
AAK96733
ID   AAK96733 standard; DNA; 1503900 BP.
XX
AC   AAK96733;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human neuregulin-1 gene.
XX
KW   Human; neuregulin 1 gene; schizophrenia; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200164877-A2.
XX
PD   07-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US06377.
XX
PR   28-FEB-2000; 2000US-0515716.
XX
PA   (DECO-) DECODE GENETICS EHF.
XX
PI   Stefansson H,  Steinthorsdottir V,  Gulcher JR;
XX
DR   WPI; 2001-514841/56.
DR   P-PSDB; AAG67938, AAG67939, AAG67940, AAG67941, AAG67942, AAG67943,
DR   AAG67944, AAG67945, AAG67946, AAG67947, AAG67948, AAG67949,
DR   AAG67950, AAG67951, AAG67952, AAG67953, AAG67954, AAG67955,
DR   AAG67956, AAG67957, AAG67958, AAG67959, AAG67960, AAG67961,
DR   AAG67962, AAG67963, AAG67964, AAG67965, AAG67966, AAG67967,
DR   AAG67968, AAG67969, AAG67970, AAG67971, AAG67972, AAG67973,
DR   AAG67974, AAG67975.
XX
PT   Neuregulin 1 nucleic acids and proteins useful for diagnosing
PT   preventing and treating schizophrenia -
XX
PS   Disclosure; Page 345-756; 756pp; English.
XX
CC   This sequence represents the human neuregulin 1 gene of the invention.
CC   The invention also relates to fragments or variants of the neuregulin 1
CC   gene. The gene and its proteins may be used in the prevention, diagnosis
CC   and treatment of diseases associated with inappropriate neuregulin 1
CC   expression, such as schizophrenia. For example they may be used to treat
CC   disorders associated with decreased neuregulin 1 expression by rectifying
CC   mutations or deletions in a patient's genome that affect the activity of
CC   neuregulin 1 by expressing inactive proteins or to supplement the
CC   patients own production of polypeptides. Additionally, the gene may be
CC   used to produce the neuregulin 1 protein, by inserting the nucleic acids
CC   into a host cell and culturing the cell to express the protein. The gene
CC   and its complementary sequences may also be used as DNA probes in
CC   diagnostic assays to detect and quantitate the presence of similar
CC   nucleic acids in samples, and therefore which patients may be in need of
CC   restorative therapy. The protein may also be used as antigens in the
CC   production of antibodies against neuregulin 1 and in assays to identify
CC   modulators of neuregulin 1 expression and activity. The antibodies and
CC   antagonists may also be used to down regulate expression and activity.
CC   The antibodies may also be used as diagnostic agents for detecting the
CC   presence of neuregulin 1 in samples.
XX
SQ   Sequence 1503900 BP; 452487 A; 281874 C; 288074 G; 480092 T; 1373 other;

  Query Match             8.1%;  Score 161.6;  DB 22;  Length 1503900;
  Best Local Similarity   82.5%;  Pred. No. 1.4e-30;
  Matches  221;  Conservative    0;  Mismatches   44;  Indels    3;  Gaps    3;

Qy        304 ACATTACCTTTTCTATGTTTAGATATG-TTAGATACACAAATATATTTCATTGTGTTATA 362
              ||  |||||||||||||||||| |||| |||||| ||||||||| |  |||| |||| ||
Db      73826 ACTGTACCTTTTCTATGTTTAGTTATGTTTAGATTCACAAATATTTCCCATTTTGTTGTA 73885

Qy        363 ATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAG 422
              ||| ||||||||||||||||||||||||||||||  ||||||| |||||||| |||||
Db      73886 ATTGCCTACAGTATTCAGTACAGTAACATGCTGTACCAGTTTGTAGCCTAGGAGCAATAG 73945

Qy        423 GCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACA 482
              || ||||||| ||| ||  | || | |||||||  || ||||||||||||||||||| ||
Db      73946 ACTCTACCATATAGCCTACATCTGCAATAGGCTA-CACTATCTAGGTTTGTGTAAGTTCA 74004

Qy        483 TTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCAC 542
              |||||||| ||| | |||||| |||||||||  |  |||||||||||||| || || |
Db      74005 CTCTATGATGTTCACCCAATGAAGAAATCACCCGA-GACACATTTCTCAGAACGTATCCC 74063

Qy        543 TGTTGTGAAGTGACCCATGACTATATTT 570
              |||  | ||||||  ||||||| || ||
Db      74064 TGTCATTAAGTGAAACATGACTGTAGTT 74091


RESULT 11
ABQ88198/c
ID   ABQ88198 standard; cDNA; 154902 BP.
XX
AC   ABQ88198;
XX
DT   18-SEP-2002  (first entry)
XX
DE   Human osteoblast differentiation related cDNA SEQ ID NO 105.
XX
KW   Human; osteoblast; stem cell differentiation; bone tissue deposition;
KW   osteoporosis; osteopathic; ss.
```

```
XX
OS   Homo sapiens.
XX
PN   WO200250301-A2.
XX
PD   27-JUN-2002.
XX
PF   18-DEC-2001; 2001WO-US48276.
XX
PR   18-DEC-2000; 2000US-255882P.
PR   24-APR-2001; 2001US-285691P.
XX
PA   (GENE-) GENE LOGIC INC.
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;
PI   Mertz L;
XX
DR   WPI; 2002-557663/59.
XX
PT   Use of genes and their expression profiles associated with osteoblast
PT   differentiation for screening modulators bone formation, for diagnosing
PT   or treating e.g. osteoporosis, or as markers for the differentiation
PT   process -
XX
PS   Claim 1; SEQ ID NO 105; 78pp + Sequence Listing; English.
XX
CC   The invention relates to genes and their expression profiles are used
CC   for:
CC   (a) screening modulators of precursor stem cell differentiation into
CC   osteoblasts, or bone tissue deposition;
CC   (b) diagnosing abnormal deposition of bone tissue, abnormal rate of
CC   osteoblast formation or osteoporosis; or
CC   (c) treating or monitoring treatment of the conditions cited in (b), or
CC   monitoring the progression of bone tissue deposition.
CC   Specific conditions include postmenopausal osteoporosis, glucocorticoid
CC   osteoporosis or male osteoporosis, osteopenia, osteodystrophy,
CC   drug-induced abnormalities in bone formation or bone loss, conditions
CC   that involve altered bone metabolism (e.g. idiopathic juvenile
CC   osteoporosis), skeletal disease linked to breast cancer, mastocytosis,
CC   Fanconi syndrome or fibrous dysplasia. The present sequence is that of an
CC   osteoblast differentiation associated cDNA marker of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 154902 BP; 43917 A; 31458 C; 32848 G; 46679 T; 0 other;

  Query Match             8.1%;  Score 161.2;  DB 24;  Length 154902;
  Best Local Similarity   80.9%;  Pred. No. 7.2e-31;
  Matches  212;  Conservative    0;  Mismatches   48;  Indels    2;  Gaps    2;

Qy        308 TACCTTTTCTATGTTTAGATATG-TTAGATACACAAATATATTTCATTGTGTTATAATTT 366
              ||||||||| || |||||||||| ||||| | ||| |||  |  |||||||||| ||||
Db     138160 TACCTTTTCAATATTTAGATATGTTTAGAAATACAGATACTTACCATTGTGTTACAATTG 138101

Qy        367 CCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAGGCTA 426
              ||||||| || ||||||||||||||||||||||||||||| ||| ||||||||  ||||
Db     138100 CCTACAGCATGCAGTACAGTAACATGCTGTACAGGTTTGTAGCCTGGGAGTAATTTGCTA 138041

Qy        427 TACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACATTCT 486
               ||||| ||| |||||||||||| ||| ||| |||||||||||| |||||||   | |||
Db     138040 CCCCATATAGCCTAGGTGTGTAGGAGGTTAT-ACCATCTAGGTTGGTGTAAGCGTACTCT 137982

Qy        487 ATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCACTGTT 546
              || || ||   |||||||  ||||| |||||| |||||||| |||||| ||  | |  ||
Db     137981 ATAATGTTTATACAATGACAAAATCTCCTAACAACACATTTTTCAGAACGTGCCTCCTTT 137922

Qy        547 GTGAAGTGACCCATGACTATAT 568
              || ||| ||| |||||||| ||
Db     137921 GTTAAGCGACACATGACTACAT 137900

RESULT 12
AAK62341/c
ID   AAK62341 standard; cDNA; 402 BP.
XX
AC   AAK62341;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:7401.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
```

PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
DR P-PSDB; AAM89560.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Claim 1; SEQ ID NO 7401; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 402 BP; 125 A; 71 C; 76 G; 125 T; 5 other;

Query Match 8.0%; Score 160; DB 22; Length 402;
Best Local Similarity 77.7%; Pred. No. 1.4e-31;
Matches 205; Conservative 1; Mismatches 56; Indels 2; Gaps 1;

```
Qy      304 ACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTATAA 363
            ||  |||||||||||| ||:|      ||  |||  ||||||  |  ||||||||  ||
Db      262 ACCGTACCTTTTCTATATTYACTATGTTTCAATATGCAAATACTTACCATTGTGT--TAG 205

Qy      364 TTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAGG 423
            || |||||||||||| |||| | | ||||||||| |||| |||| |||||||| ||||||
Db      204 TTGCCTACAGTATTCTGTACGGCATCATGCTGTAGAGGTCTGTAGCCTAGGAGCAATAGG 145

Qy      424 CTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACAT 483
            ||||||||||||||  |||||||||| ||||| || | |||| ||||||||||||||||
Db      144 CTATACCATACAGCCCAGGTGTGTAGCAGGCTTTACCAATCTCGGTTTGTGTAAGTACAC 85

Qy      484 TCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCACT 543
            || ||||| ||  ||||||||||||||   |||||| |||  ||| ||||||| | || |
Db       84 TCCATGATGTTTGCACAATGATGAAAGTGCCTAACAACATGTTTGTCAGAATATATCCCC 25

Qy      544 GTTGTGAAGTGACCCATGACTATA 567
             |||| || |||   |||||||||
Db       24 ATTGTTAATTGATATATGACTATA 1
```

RESULT 13
AAC79685/c
ID AAC79685 standard; cDNA; 1332 BP.

```
XX
AC   AAC79685;
XX
DT   12-FEB-2001  (first entry)
XX
DE   Human secreted protein gene 5 SEQ ID NO:15.
XX
KW   Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; gene therapy;
KW   food additive; preservative; chromosome identification; cancer;
KW   immune disorder; cardiovascular disorder; neurological disease;
KW   wound healing; infectious disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058339-A2.
XX
PD   05-OCT-2000.
XX
PF   22-MAR-2000; 2000WO-US07440.
XX
PR   26-MAR-1999;   99US-0126503.
PR   17-DEC-1999;   99US-0172409.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis G;
XX
DR   WPI; 2000-594637/56.
DR   P-PSDB; AAB44600.
XX
PT   Fifty nucleic acid molecules encoding human secreted proteins, useful
PT   in the prevention, treatment and diagnosis of cancer, immune disorders,
PT   cardiovascular disorders and neurological diseases -
XX
PS   Claim 1; Page 339-340; 410pp; English.
XX
CC   The polynucleotide sequences given in AAC79681 to AAC79730 encode the
CC   human secreted proteins given in AAB44596 to AAB44645. AAB44646 to
CC   AAB44693 represent human secreted polypeptide sequences and proteins
CC   homologous to them, which are given in the exemplification of the present
CC   invention. Human secreted proteins have activities based on the tissues
CC   and cells the genes are expressed in. Examples of activities include:
CC   cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
CC   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
CC   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic; and
CC   cardiant. The polynucleotides and polypeptides are useful for preventing,
CC   treating or ameliorating a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. The polypeptides
CC   can also be used as a food additive or preservative to increase or
CC   decrease storage capabilities. The polynucleotides are useful for
CC   chromosome identification. They are also useful as probes for diagnosing
CC   a disorder related to the female reproductive system, particularly breast
CC   and/or ovary cancer. They are also useful in the gene therapy of breast
CC   and ovarian cancer. The nucleic acids, protein, antibodies, agonists and
CC   antagonists from the present invention are useful in the diagnosis,
CC   treatment and prevention of: cancer; immune disorders; cardiovascular
CC   disorders; wound healing; neurological diseases; and infectious
CC   diseases. AAC79672 to AAC79680 and AAB44595 represent sequences used in
CC   the exemplification of the present invention.
XX
SQ   Sequence 1332 BP; 446 A; 270 C; 259 G; 357 T; 0 other;

  Query Match             7.9%;  Score 157.8;  DB 21;  Length 1332;
  Best Local Similarity   78.4%;  Pred. No. 8.6e-31;
  Matches  239;  Conservative    0;  Mismatches   62;  Indels    4;  Gaps    4;

Qy        287 CATAAGTGTAATGCAGGACATTACCTTTTCTATGTTTAGATATG-TTAGATACACAAATA 345
              |||||  ||     | |  | ||||||| |||| |||| ||| | ||||||||||||||
Db        920 CATAAAAGTTTACTACGGTAGTACCTTTGCTATATTTAAATAGGTTTAGATACACAAATA 861

Qy        346 TATTTCATTGTGTTATAATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTG 405
              |  ||| |||||||||||| |||||||||||||||||| |||||||||| ||||||||
Db        860 CTTGCCATAGTGTTATAATTGCCTACAGTATTCAGTACA-TAACATGCTGTGCAGGTTTG 802

Qy        406 TAACCTAGGAGTAATAGGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCT 465
              || |||||||| | |||| ||||||||| ||  |||||||| |||||| ||| |||||||
Db        801 TAGCCTAGGAGCACTAGGTTATACCATATAGGATAGGTGTGCAGTAGGTTAT-ACCATCT 743

Qy        466 AGGTTTGTGTAAGTACATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACAT 525
              || ||||||||||||  ||||| ||| ||||||  | || ||||||||||| ||||||
Db        742 AGATTTGTGTAAGTATGTTCTACGATGTTCCCATCA-GAAGAAATCACCTAATGACACAT 684

Qy        526 TTCTCAGAATGTTTCACTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGAT 585
              |||||||||||| || |  |  | ||||||| | ||| |  | |||| | | |  || |
Db        683 TTCTCAGAATGTATCCCCATCATTAAGTGACACTTGATTGCAATTTCTTTTTTTTTTTT 624

Qy        586 ATTTT 590
              |||||
Db        623 TTTTT 619


RESULT 14
AAH92899
ID   AAH92899 standard; DNA; 700 BP.
XX
AC   AAH92899;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Human inflammatory bowel disease related gene fragment IGR1365a.
XX
KW   Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;
KW   single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;
KW   chromosome 5q31-33; forensic test; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200142511-A2.
XX
PD   14-JUN-2001.
XX
PF   11-DEC-2000; 2000WO-US33632.
XX
PR   10-DEC-1999;   99US-0170257.
PR   10-APR-2000; 2000US-0196046.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA   (ELLI-) ELLIPSIS BIOTHERAPEUTICS CORP.
XX
PI   Daly M,  Hudson TJ,  Lander ES,  Rioux J,  Siminovitch K;
XX
DR   WPI; 2001-367874/38.
XX
PT   Testing for the presence of polymorphisms associated with inflammatory
PT   bowel disease, using a hybridization assay -
XX
PS   Disclosure; Page 347; 463pp; English.
XX
CC   The present invention describes a method for detecting the presence of
CC   polymorphisms associated with inflammatory bowel diseases such as
CC   ulcerative colitis and Crohn's disease. The methods can be used to detect
CC   the presence of genetic polymorphisms associated with inflammatory bowel
CC   disease and correlating their occurrence with disease states. They may be
CC   used in this way for phenotypic correlations, forensics, paternity
CC   testing, medicine and genetic analysis. The present sequence is a gene
CC   containing a polymorphic site described in the exemplification of the
CC   invention.
XX
SQ   Sequence 700 BP; 213 A; 143 C; 142 G; 202 T; 0 other;

  Query Match             7.8%;  Score 155.8;  DB 22;  Length 700;
  Best Local Similarity   79.5%;  Pred. No. 2.2e-30;
  Matches  209;  Conservative    0;  Mismatches   52;  Indels    2;  Gaps    2;
```

```
Qy        304 ACATTACCTTTTCTATGTTTAGATATG-TTAGATACACAAATATATTTCATTGTGTTATA 362
              ||  | |||||||| |||||||||||| |||||||||||||||| |  | ||||||  |
Db        439 ACTGTGCCTTTTCTGTGTTTAGATATGCTTAGATACACAAATATTTACCCTTGTGTGGCA 498

Qy        363 ATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAG 422
               |  ||||||||  |||| | ||| || |||||||||| ||||| |||||||| |||||
Db        499 GTCGCCTACAGTGCTCAGCAGAGTTACTTGCTGTACAGGCTTGTACCCTAGGAGCAATAG 558

Qy        423 GCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACA 482
              |||||||||| | ||| ||||||| | ||||| ||| ||||||||||||||||||||||||
Db        559 GCTATACCACATAGCCTAGGTGTTTGGTAGGTTAT-ACCATCTAGGTTTGTGTAAGTACA 617

Qy        483 TTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCAC 542
              ||||||||||| ||||| ||  |||| ||||||  |  || |||||||| ||| ||
Db        618 CTCTATGATATTCACACAAGGACAAAATTACCTAATGAAGCACTTCTCAGACTGTATCCT 677

Qy        543 TGTTGTGAAGTGACCCATGACTA 565
              ||||   |||  |  ||||| ||
Db        678 TGTTACTAAGCAATACATGATTA 700
```

```
RESULT 15
AAX35699/c
ID   AAX35699 standard; cDNA; 2604 BP.
XX
AC   AAX35699;
XX
DT   09-JUL-1999  (first entry)
XX
DE   cDNA encoding a protein identified by the signal sequence trap method.
XX
KW   Signal sequence trap method; SST method; immunisation; inhibition;
KW   infection; allergy; cancer; regulation; tissue formation; tissue repair;
KW   activin activity; inhibin activity; chemokine activity;
KW   cytokine activity; blood coagulation regulation; agonist; antagonist;
KW   metabolic disorder; hormonal disorder; immune disorder;
KW   severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer;
KW   wound; ss.
XX
OS   Homo sapiens.
XX
PN   WO9918126-A1.
XX
PD   15-APR-1999.
XX
PF   06-OCT-1998;   98WO-JP04514.
XX
PR   07-OCT-1997;   97JP-0274674.
XX
PA   (ONOY ) ONO PHARM CO LTD.
XX
PI   Fukushima D,  Shibayama S,  Tada H;
XX
DR   WPI; 1999-277254/23.
DR   P-PSDB; AAY02360.
XX
PT   Polypeptides identified by the signal sequence trap method from a
PT   human cDNA library
XX
PS   Claim 5; Page 75-77; 281pp; Japanese.
XX
CC   AAX35694-X35747 represent cDNA sequences that encode novel polypeptides
CC   (AAY02358-84) which are identified from a human placental cDNA library
CC   by the signal sequence trap (SST) method. The polypeptides have a
CC   broad range of physiological activity, including immunisation against
CC   and inhibition of infections, allergies and cancer; regulation of tissue
CC   formation and repair; activin/inhibin activity; chemokine/cytokine
CC   activity; blood coagulation regulation; and receptor/ligand agonist
CC   or antagonist activity. The polypeptides can be used for prevention
CC   and treatment of disorders including infections by bacteria, yeasts and
CC   viruses (including HIV) and protozoa; metabolic and hormonal disorders;
CC   immune disorders (including severe combined immunodeficiency (SCID)
CC   and AIDS; thrombosis; cancer; and traumatic or surgical wounds.
XX
SQ   Sequence 2604 BP; 779 A; 556 C; 607 G; 662 T; 0 other;
```

```
  Query Match             7.8%;  Score 155.6;  DB 20;  Length 2604;
  Best Local Similarity   71.6%;  Pred. No. 4.2e-30;
  Matches  232;  Conservative    0;  Mismatches   89;  Indels    3;  Gaps    2;

Qy        304 ACATTACCTTTTCTATGTTTAGATATG-TTAGATACACAAATATATTTCATTGTGTTATA 362
              || || | |||||| |||||| ||||| ||||||||||||||| |  |||||| | |
Db       1511 ACTTTCCTTTTTCTCTGTTTACATATGTTTAGATACACAAATATTTCCCATTGTGCTGCA 1452

Qy        363 ATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAG 422
              ||| |||| ||||||||||  ||| ||||||||||||||  ||| ||| |||| |||||
Db       1451 ATTGCCTATAGTATTCAGT--AGTCACATGCTGTACAGGTGCGTAGCCTGGGAGCAATAG 1394

Qy        423 GCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACA 482
              ||| | ||||| ||| |||||| || |||||||||| |     || ||||| ||||||||
Db       1393 GCTCTCCCATATAGCCTAGGTATGCAGTAGGCTATGCCATCTAAGTTTTGTATAAGTACA 1334

Qy        483 TTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCAC 542
              ||| ||||| || ||| |  ||| |||  |||||  |||||||| ||||||||  || |
Db       1333 ATCTGTGATAATCACACTACAATGCAATTGCCTAATGACACATTTATCAGAATGCGTCCC 1274

Qy        543 TGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGCC 602
              |||    ||| ||  ||||||| |||| | ||   |  ||  | ||||| |   | |
Db       1273 TGTCACCAAGAGATACATGACTGTATTCTTCTTCTTTTTTTTTTTTTTTTTTTTTTGA 1214

Qy        603 CATGAGAATGTAGTGTAAGATCAA 626
              |||| |     | | | | | | |
Db       1213 GATGGGGTCTCATTCTGACACCCA 1190
```

```
Search completed: November  7, 2003, 06:15:39
Job time : 629.98 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 6, 2003, 21:13:51 ; Search time 155.728 Seconds
(without alignments)
8684.478 Million cell updates/sec

Title: US-09-939-209A-3_COPY_10000_10500
Perfect score: 501
Sequence: 1 gaccatgtataatatgatgc.....cgctcaaagccgaagccaca 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
✓	1	501	100.0	20300	24	ABK47337	Genomic nucleotide
	2	180.6	36.0	1238	22	AAL26766	Human breast cance
	3	177.4	35.4	3165	25	ABX63113	Human cDNA #113 di
	4	45.4	9.1	10020	24	ABL34293	Human immune syste
	5	41.2	8.2	8845	22	AAS46544	Tumour suppressor
c	6	40.6	8.1	1456	24	ABQ61218	FLJ23018 fis clone
	7	40.2	8.0	5661	24	ABQ67088	Human angiogenesis
	8	40	8.0	306	24	ABN94869	Gene #1367 used to
	9	40	8.0	306	24	ABL63962	Breast cancer rela
	10	39.8	7.9	9646	24	ABL33688	Human immune syste
	11	39.6	7.9	2360	21	AAC66096	pSS3 cDNA sequence
	12	39.6	7.9	2494	21	AAC66090	Rice sucrose synth
c	13	38.6	7.7	648	22	AAK69902	Human immune/haema
	14	38.2	7.6	2857	19	AAV54123	Rabbit membrane pr
	15	38.2	7.6	15732	22	AAS45388	Chemically pretrea
	16	38.2	7.6	15732	24	ABK28233	DNA transcription
c	17	38	7.6	140167	24	ABT10146	Human breast cance
c	18	37.8	7.5	861	20	AAX99566	Nucleic acid seque
c	19	37.8	7.5	8372	23	ABL23536	Drosophila melanog
	20	37.8	7.5	11473	24	ABK40030	Human chemically p
	21	37.8	7.5	11473	24	ABL33355	Human immune syste
c	22	37.2	7.4	513	24	ABQ76618	C. albicans BAX-as
	23	37.2	7.4	6228	24	ABL70469	Chemically treated
	24	37.2	7.4	6228	24	AAS61430	Human gene regulat
	25	37.2	7.4	6228	24	ABK31496	Signal transductio
	26	37.2	7.4	112132	24	ABK90888	Human ATP-dependen
c	27	37	7.4	7508	24	ABK31206	Signal transductio
	28	37	7.4	9964	24	ABL32099	Human immune syste
	29	37	7.4	10048	24	ABQ67015	Human angiogenesis
	30	36.8	7.3	543	22	AAH53661	S. epidermidis ope
c	31	36.8	7.3	3666	22	AAH54866	S. epidermidis gen
c	32	36.6	7.3	487	22	AAK18805	Human brain expres
c	33	36.6	7.3	487	22	AAK44744	Human bone marrow
c	34	36.6	7.3	487	22	AAI50721	Probe #19407 used
c	35	36.6	7.3	487	23	ABS44404	Human liver single
c	36	36.6	7.3	487	24	ABS18983	Human genome-deriv
c	37	36.6	7.3	491	22	AAK06024	Human brain expres
c	38	36.6	7.3	491	22	AAK31670	Human bone marrow
c	39	36.6	7.3	491	22	AAI37546	Probe #6232 used t
c	40	36.6	7.3	491	23	ABS31354	Human liver single
c	41	36.6	7.3	491	24	ABS06426	Human genome-deriv
	42	36.6	7.3	853	21	AAA81913	N. meningitidis pa
	43	36.6	7.3	978	21	AAC64462	*Arabidopsis profil*
	44	36.6	7.3	1162	21	AAC64461	Arabidopsis profil
	45	36.6	7.3	1388	21	AAC64460	Arabidopsis profil

ALIGNMENTS

RESULT 1
ABK47337
ID ABK47337 standard; DNA; 20300 BP.
XX
AC ABK47337;
XX
DT 18-JUN-2002 (first entry)
XX
DE Genomic nucleotide sequence encoding human RGS-4 protein.
XX
KW RGS-4; schizophrenia; human; regulator of G protein signalling 4;
KW neuroleptic; polymorphism; schizotypy; schizoid; chromosome 1q21-22;
KW gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace (4121,T)
FT /*tag= a
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4123,A)
FT /*tag= b
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4368,C)
FT /*tag= c
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4621,C)
FT /*tag= d
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4790,T)
FT /*tag= e

```
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (4816,T)
FT                   /*tag= f
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (4970,T)
FT                   /*tag= g
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace 5055,G)
FT                   /*tag= h
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (5295,A)
FT                   /*tag= i
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (5695,A)
FT                   /*tag= j
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (7375,T)
FT                   /*tag= k
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (7759,A)
FT                   /*tag= l
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (8596,A)
FT                   /*tag= m
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (9602..9610,CA)
FT                   /*tag= n
FT                   /note= "Deletion polymorphism"
FT   variation       replace (9892,A)
FT                   /*tag= o
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (9963,A)
FT                   /*tag= p
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (10132,A)
FT                   /*tag= q
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11056,C)
FT                   /*tag= r
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11091,T)
FT                   /*tag= s
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11106,A)
FT                   /*tag= t
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11774,T)
FT                   /*tag= u
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (12143,A)
FT                   /*tag= v
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (12145,T)
FT                   /*tag= w
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (14367,G)
FT                   /*tag= x
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (17027..17029,GA)
FT                   /*tag= y
FT                   /note= "Deletion polymorphism"
FT   variation       replace (17630,T)
FT                   /*tag= Z
FT                   /standard_name= "Single-nucleotide polymorphism"
XX
PN   WO200216653-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26622.
XX
PR   24-AUG-2000; 2000US-228021P.
XX
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Levitt PR, Mirnics K, Kodavali VC, Nimgaonkar VL;
XX
DR   WPI; 2002-292070/33.
XX
PT   Diagnosing, assessing susceptibility and treating schizophrenia,
PT   involves observing regulator of G-protein signalling 4, RGS4 levels in a
PT   subject -
XX
PS   Claim 1; Page 20-33; 112pp; English.
XX
CC   This invention relates to a novel method for diagnosing schizophrenia
CC   or determining susceptibility to schizophrenia in a human. The method
CC   comprises obtaining from a patient a DNA sample and detecting variations
CC   in the regulator of G-protein signalling 4 (RGS4) gene. Alternatively,
CC   the method involves measuring RGS4 mRNA or protein levels in a tissue
CC   sample from the patient and determining if there is a reduced level.
CC   The method of the invention is useful for diagnosing and determining
CC   susceptibility to schizophrenia. The invention also comprises a method
CC   that is useful for treating schizophrenia which includes a prophylactic
CC   treatment. The method of genotyping polymorphic variants in the RGS-4
CC   gene is applied to diagnosing pathologies of the schizophrenic spectrum,
CC   such as in particular schizotypy, schizoid individuals, etc. This
CC   method offers the possibility of diagnosing schizophrenia by a
CC   biological test and no longer exclusively by clinical evaluations.
CC   The present sequence represents the genomic DNA encoding the human
CC   regulator of G-protein signalling 4 (RGS4) protein used in the method of
CC   the invention. The gene for the RGS4 protein is located on human
CC   chromosome 1q21-22.
XX
SQ   Sequence 20300 BP; 6157 A; 4102 C; 3775 G; 6266 T; 0 other;

  Query Match            100.0%;  Score 501;  DB 24;  Length 20300;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-132;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GACCATGTATAATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10000 GACCATGTATAATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCT 10059

Qy         61 AAGTAAGCTCCAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10060 AAGTAAGCTCCAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATT 10119

Qy        121 TTTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10120 TTTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCT 10179

Qy        181 TTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10180 TTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATAT 10239

Qy        241 CCCTACTTTTCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10240 CCCTACTTTTCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCT 10299

Qy        301 TTTCTTCCTCATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10300 TTTCTTCCTCATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTG 10359

Qy        361 CAGAGCGGTCGTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10360 CAGAGCGGTCGTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAG 10419

Qy        421 GCTTAGCAGGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10420 GCTTAGCAGGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGT 10479

Qy        481 ACGCTCAAAGCCGAAGCCACA 501
              |||||||||||||||||||||
Db      10480 ACGCTCAAAGCCGAAGCCACA 10500
```

```
RESULT 2
AAL26766
ID   AAL26766 standard; cDNA; 1238 BP.
XX
AC   AAL26766;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 19223.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer  -
XX
PS   Claim 1; Page 3683-3684; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 1238 BP; 386 A; 265 C; 307 G; 274 T; 6 other;

  Query Match             36.0%;  Score 180.6;  DB 22;  Length 1238;
  Best Local Similarity   97.9%;  Pred. No. 2.2e-41;
  Matches  183;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        315 CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 374
              ||    | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         58 CTCCTAAGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 117

Qy        375 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        118 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 177

Qy        435 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        178 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 237

Qy        495 AGCCACA 501
              |||||||
Db        238 AGCCACA 244

RESULT 3
ABX63113
ID   ABX63113 standard; cDNA; 3165 BP.
XX
AC   ABX63113;
XX
DT   25-FEB-2003  (first entry)
XX
DE   Human cDNA #113 differentially expressed in activated vascular tissue.
XX
KW   Human; gene; ss; vascular tissue; cytostatic; atherosclerosis;
KW   cardiant; hypotensive; antidiabetic; gynaecological; vasotropic;
KW   cerebroprotective; gene therapy; vascular disease; cancer; coronary;
KW   artery disease; hypertension; diabetes; pre-eclampsia; restenosis;
KW   ischaemia-reperfusion injury; stroke;
XX
OS   Homo sapiens.
XX
PN   US2002137081-A1.
XX
PD   26-SEP-2002.
XX
PF   08-JAN-2002; 2002US-0044090.
XX
PR   28-JUL-2000; 2000US-222469P.
PR   08-JAN-2001; 2001US-260483P.
XX
PA   (BAND/) BANDMAN O.
XX
PI   Bandman O;
XX
DR   WPI; 2003-110597/10.
XX
PT   Combination for diagnosing, staging, treating, or monitoring the
PT   progression of treatment of a vascular disease, e.g. atherosclerosis,
PT   comprises several cDNAs that are differentially expressed in activated
PT   vascular tissue -
XX
PS   Claim 1; Page -; 18pp; English.
XX
CC   This invention relates to a combination comprising several cDNAs that
CC   are differentially expressed in activated vascular tissue. The invention
CC   also discloses a high throughput method for detecting differentially
CC   expressed cDNAs in a sample. The cDNAs of the invention may have
CC   antiarteriosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;
CC   gynaecological; vasotropic and cerebroprotective activities and may be
CC   used in gene therapy. The cDNAs of the invention may be used in a
CC   high-throughput methods for detecting differential expression of one or
CC   more cDNAs in a sample, or screening several molecules or compounds to
CC   identify a molecule or compound that specifically binds a cDNA of the
CC   invention. A protein encoded by the cDNA may be used to screen several
CC   molecules or compounds to identify a ligand that specifically binds to
CC   the protein, or to produce or purify an antibody to the protein that can
CC   be used to detect a protein in a sample or purify a natural or
CC   recombinant protein from a sample. The nucleotides may be useful for
CC   diagnosing, staging, treating, or monitoring the progression of
CC   treatment of a vascular disease, e.g. atherosclerosis, cancer, coronary
CC   artery disease, hypertension, diabetes, pre-eclampsia, ischaemia-
CC   reperfusion injury, restenosis, or stroke. The cDNAs can also be used
CC   for large-scale genetic or gene expression analysis of several new
CC   nucleic acid molecules. Antibodies to the proteins encoded by the
CC   cDNAs are useful for diagnosing pre-pathologic disorders, and chronic
CC   or acute diseases associated with abnormalities in the expression,
CC   amount or distribution of the protein. The present sequence
CC   represents a cDNA of the invention that is differentially expressed in
CC   activated vascular tissue.
CC   Note: The sequence data for this patent did not form part of the
CC   specification, but was obtained in electronic format directly from USPTO
CC   at http.seqdata.uspto.gov/sequence.html?DocID=20020137081.
XX
SQ   Sequence 3165 BP; 889 A; 636 C; 672 G; 967 T; 1 other;

  Query Match             35.4%;  Score 177.4;  DB 25;  Length 3165;
  Best Local Similarity   96.8%;  Pred. No. 2.7e-40;
```

```
  Matches  181;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy        315 CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 374
              ||  ||||    ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 CGTAGAGGGCTCGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 122

Qy        375 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 182

Qy        435 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 242

Qy        495 AGCCACA 501
              |||||||
Db        243 AGCCACA 249


RESULT 4
ABL34293
ID   ABL34293 standard; DNA; 10020 BP.
XX
AC   ABL34293;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2266.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 2266; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 10020 BP; 3111 A; 92 C; 2113 G; 4704 T; 0 other;

  Query Match             9.1%;  Score 45.4;  DB 24;  Length 10020;

  Best Local Similarity   51.0%;  Pred. No. 0.024;
  Matches  132;  Conservative    0;  Mismatches  126;  Indels    1;  Gaps    1;

Qy         90 TTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTTACAGGCATATGAATAAAAACT 149
              |||  ||    |||  |  |    |    | |||||||||   ||  ||||  ||||  | |
Db       1329 TTTGTTTTAGGGATGAATTTAGTGTTCGAGTTTTTTTTTTTAGTGATATATATAAGGAAT 1388

Qy        150 ATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAGAAATTCTAAACCTCTGACATTG 209
              | | |||    ||         ||||||   ||| |   || | ||||  | ||   |||
Db       1389 TTTTATTGGTTGATATGTTTAGTTTTTTTTATTTTTGTGAAATATAAATTTTTGTTTTTG 1448

Qy        210 GTGGAGACATTGAGTACATTTTTTCCCATATCCCTACTTTTCAGAAGGATTTTCTCTGCT 269
              | | | |  | |||   ||||   || |   |  ||||  |     |||| |    |
Db       1449 ATTTA-AGAAAGGTTACGCGTTTTTTTATTTTTTTGTTTTTTTGTTTTTTTTTTTTATTT 1507

Qy        270 CGTTCACTTAACATTGCTGATGCGTCAGTCTTTTCTTCCTCATCTCTTTCAGGGGCTGGA 329
              |   || |  ||  | |   |   | ||||  |  | | |  ||| |     ||||
Db       1508 TTATTTTTTTATTTTTTTTTTTTTTTATTTTTTTATAATAAACGTTTTAAAATTATGGA 1567

Qy        330 GAGGCAGAGGGAGACAGAG 348
              |||  |||| |||| ||||
Db       1568 GAGAGAGAGAGAGAGAGAG 1586


RESULT 5
AAS46544
ID   AAS46544 standard; DNA; 8845 BP.
XX
AC   AAS46544;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #266.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW   cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02955.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumour suppressor
PT   genes and oncogenes, useful in designing primers and probes for
PT   analysing diseases associated with cytosine methylation state e.g.
PT   cancer  -
XX
PS   Claim 1; SEQ ID No 266; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and
CC   oncogenes having a sequence taken from 536 (actually 533 since
CC   numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC   (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC   peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC   form part of a set of probes for detecting the cytosine methylation state
```

```
CC   and/or single nucleotide polymorphisms and also to be used in an
CC   array for analysing diseases associated with CpG dinucleotides e.g.
CC   cancers and tumours. The probes can also be used in a method for
CC   ascertaining genetic and/or epigenetic parameters for the diagnosis
CC   and/or therapy of existing diseases or the predisposition to specific
CC   diseases, by analysing cytosine methylations. The parameters may be
CC   compared to another set of genetic and/or epigenetic parameters, the
CC   differences serving as basis for diagnosis and/or prognosis events which
CC   are disadvantageous to patients. The present sequence is one of the
CC   533 genomic sequences derived from tumour suppressor genes and
CC   oncogenes. Sequences with even numbered Seq ID numbers are the
CC   complementary sequence of the corresponding odd numbered sequence (e.g.
CC   ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC   is missing).
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 8845 BP; 2565 A; 135 C; 1694 G; 4451 T; 0 other;

  Query Match             8.2%;  Score 41.2;  DB 22;  Length 8845;
  Best Local Similarity   51.1%;  Pred. No. 0.36;
  Matches   97;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy        113 TTGATATTTTTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTT 172
              || |   |||||||||| |     || | ||  |  | || |||| ||   |  | |
Db       1719 TTAAAGGTTTTTTTTTAAAATTTAATTATTATTATGTTTAATTTGGAGTTAAGGTTATAT 1778

Qy        173 TTTTTCCTTTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTT 232
              || ||   ||| ||  | || |  |  | | | ||  ||    | ||  |  |  |||||
Db       1779 TTATTTTGTTTTTATTATTTTTTTAGTTATAAGATATGTATGAATATGTAAAATTTTTTT 1838

Qy        233 TCCCATATCCCTACTTTTCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGC 292
              |   |||     |  | || || || ||| |       | | |  | ||| |||    | |
Db       1839 TAAAATAAGATTTTTATTTAGTAGTATTATTATAATTTGATTTCGTAAAATTTTAAAGGA 1898

Qy        293 GTCAGTCTTT 302
              ||   | |||
Db       1899 GTAGTTTTTT 1908


RESULT 6
ABQ61218/c
ID   ABQ61218 standard; cDNA; 1456 BP.
XX
AC   ABQ61218;
XX
DT   26-FEB-2003  (first entry)
XX
DE   FLJ23018 fis clone LNG00903 encoding sequence.
XX
KW   Neuroprotective; immunomodulator; cancer; chromosome X;
KW   cytostatic; anti-inflammatory; gene therapy; nutritional supplement;
KW   wound; burn; ulcer; Alzheimer's disease; Huntington's disease;
KW   amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
KW   vulnerary; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200231111-A2.
XX
PD   18-APR-2002.
XX
PF   11-OCT-2001; 2001WO-US27760.
XX
PR   12-OCT-2000; 2000US-0687527.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Drmanac RT;
XX
DR   WPI; 2002-426278/45.
DR   N-PSDB; ABP43974.
XX
PT   New polypeptides and their encoded proteins, useful as nutritional
PT   sources or supplements, or in gene therapy, particularly for treating
PT   wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
PT   inflammation  -
XX
PS   Claim 1; SEQ ID # 431; 357pp + sequence listing; English.
XX
CC   The invention relates to 446 newly isolated polynucleotide sequences.
CC   The activity of polynucleotides of the invention may be described as,
CC   vulnerary, neuroprotective, immunomodulator, cytostatic and
CC   anti-inflammatory. Compositions comprising nucleic acids of the invention
CC   are useful for treating a mammalian subject, or as nutritional sources or
CC   supplements. These are useful in gene therapy, particularly for treating
CC   wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
CC   amyotrophic lateral sclerosis, autoimmune disorders, cancer or
CC   inflammation. The nucleic acids and polypeptides are also useful in
CC   diagnostic and research methods. The sequences given in records
CC   ABQ60788-ABQ61233 represent polynucleotides of the invention.
CC   NOTE: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1456 BP; 377 A; 392 C; 302 G; 385 T; 0 other;

  Query Match             8.1%;  Score 40.6;  DB 24;  Length 1456;
  Best Local Similarity   52.0%;  Pred. No. 0.25;
  Matches   91;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy        284 TGCTGATGCGTCAGTCTTTTCTTCCTCATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGA 343
              || |  |  ||  |   | ||| || ||||     |||| |  ||||  | ||  ||||
Db        958 TGGTTGTAAGTTTGAAGTCTCTCCCACATCAAGAGCAGGAGGAGGAGGAGGAGGAGGAGG 899

Qy        344 CAGAGGAGCTGGTACTGCAGAGCGGTCGTCTGATTGGCTGGACGGTCGTAGCTGGGCTAT 403
              ||||||  |||  ||  |   |       || | |    |  ||| || | || | ||
Db        898 AGGAGGAGGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTTGTGGTAA 839

Qy        404 AAAAGAGACCCCTACAGGCTTAGCAGGAAGACGCTCAGAGGATTCTGACAATATC 458
              | | |   |  |   |||  |||||| | | |    | |||||  |||    |||
Db        838 AGAGGCACCAGCATGAGGAATAGCAGAATGGCATGGATAGGATGGTGAGGCAATC 784


RESULT 7
ABQ67088
ID   ABQ67088 standard; DNA; 5661 BP.
XX
AC   ABQ67088;
XX
DT   28-AUG-2002  (first entry)
XX
DE   Human angiogenesis associated polynucleotide SEQ ID NO 118.
XX
KW   Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW   inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;
KW   macular degeneration; inflammatory bowel disease; Crohn's disease;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiarteriosclerotic; ds.
XX
OS   Homo sapiens.
XX
PN   WO200246454-A2.
XX
PD   13-JUN-2002.
XX
PF   06-DEC-2001; 2001WO-EP14320.
XX
PR   06-DEC-2000; 2000DE-1061338.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Schacht O;
```

XX
DR WPI; 2002-500450/53.
XX
PT New nucleic acid fragments from chemically treated
PT angiogenesis-associated genes, useful for determining methylation
PT status, e.g. in diagnosis or treatment of cancer -
XX
PS Claim 1; SEQ ID NO 118; 41pp + Sequence Listing; German.
XX
CC The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5661 BP; 1562 A; 46 C; 1104 G; 2945 T; 4 other;

Query Match 8.0%; Score 40.2; DB 24; Length 5661;
Best Local Similarity 51.4%; Pred. No. 0.58;
Matches 93; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 107 ACTTCCTTGATATTTTTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATTGT 166
Db 2028 ATTTATGTTTTATTTTTTTTTTTTGGAAGTATTTATTGAAAATATATAGTGGAGAAATAT 2087

Qy 167 ACACTTTTTTTCCTTTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTAC 226
Db 2088 TGATTTGTTTTATTATTTTAGTATTTTATGGTGTTTTTTATTTTTTTTATTGTATTAGGT 2147

Qy 227 ATTTTTTCCCATATCCCTACTTTTCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGC 286
Db 2148 TTGTTTTTTAATATTTTTTTTTTTTTATGGATTTTTGTTTTTTTTTTGCGTTAAGATTGA 2207

Qy 287 T 287
Db 2208 T 2208

RESULT 8
ABN94869
ID ABN94869 standard; DNA; 306 BP.
XX
AC ABN94869;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #1367 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30589.
XX
PR 02-OCT-2000; 2000US-237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

XX
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
PT liver tissue sample -
XX
PS Claim 1; SEQ ID NO 1367; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 306 BP; 98 A; 55 C; 40 G; 113 T; 0 other;

Query Match 8.0%; Score 40; DB 24; Length 306;
Best Local Similarity 59.8%; Pred. No. 0.19;
Matches 67; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 119 TTTTTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTC 178
Db 1 TTTTTTTTTCAGTGACTTATCAAAAATTTATTTCATATAATAAATTATATAATTTATTTT 60

Qy 179 CTTTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTT 230
Db 61 CATCTTTAAACAGTCTACACCGAAAACATTTTTGGAAACATCTTTTCCTTTT 112

RESULT 9
ABL63962
ID ABL63962 standard; DNA; 306 BP.
XX
AC ABL63962;
XX
DT 15-MAY-2002 (first entry)
XX
DE Breast cancer related gene sequence SEQ ID NO:2299.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.

```
PR   25-SEP-2000; 2000US-234923P.
PR   25-SEP-2000; 2000US-234924P.
PR   25-SEP-2000; 2000US-235077P.
PR   25-SEP-2000; 2000US-235082P.
PR   25-SEP-2000; 2000US-235134P.
PR   25-SEP-2000; 2000US-235280P.
PR   26-SEP-2000; 2000US-235637P.
PR   26-SEP-2000; 2000US-235638P.
PR   27-SEP-2000; 2000US-235711P.
PR   27-SEP-2000; 2000US-235720P.
PR   27-SEP-2000; 2000US-235840P.
PR   27-SEP-2000; 2000US-235863P.
PR   28-SEP-2000; 2000US-236028P.
PR   28-SEP-2000; 2000US-236032P.
PR   28-SEP-2000; 2000US-236033P.
PR   28-SEP-2000; 2000US-236034P.
PR   28-SEP-2000; 2000US-236109P.
PR   28-SEP-2000; 2000US-236111P.
PR   29-SEP-2000; 2000US-236842P.
PR   29-SEP-2000; 2000US-236891P.
PR   02-OCT-2000; 2000US-237172P.
PR   02-OCT-2000; 2000US-237173P.
PR   02-OCT-2000; 2000US-237278P.
PR   02-OCT-2000; 2000US-237294P.
PR   02-OCT-2000; 2000US-237295P.
PR   02-OCT-2000; 2000US-237316P.
PR   03-OCT-2000; 2000US-237425P.
PR   03-OCT-2000; 2000US-237598P.
PR   03-OCT-2000; 2000US-237604P.
PR   03-OCT-2000; 2000US-237606P.
PR   03-OCT-2000; 2000US-237608P.
PR   01-NOV-2000; 2000US-244867P.
PR   01-NOV-2000; 2000US-245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE,  Augustus M,  Carter KC,  Ebner R,  Endress G,  Horrigan S;
PI   Soppet DR,  Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a
PT   chemical agent to be tested for anti-neoplastic activity, and
PT   determining a change in expression of a gene of a signature gene set  -
XX
PS   Claim 1; SEQ ID 2299; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an
CC   anti-neoplastic agent. The method involves exposing cells to a chemical
CC   agent to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening
CC   an anti-neoplastic agent, and can be used for producing a product which
CC   is the data collected with respect to the anti-neoplastic agent as a
CC   result of M1, and the data is sufficient to convey the chemical
CC   structure and/or properties of the agent. M1 can be used in the
CC   treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC   oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC   adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC   infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC   carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ   Sequence 306 BP; 98 A; 55 C; 40 G; 113 T; 0 other;

  Query Match             8.0%;  Score 40;  DB 24;  Length 306;
  Best Local Similarity   59.8%;  Pred. No. 0.19;
  Matches   67;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy        119 TTTTTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTC 178
              |||||||||| |  | | ||| || ||    | |  | |     ||| || | ||| |||

Db          1 TTTTTTTTTCAGTGACTTATCAAAAATTTATTTCATATAATAAATTATATAATTTATTTT 60

Qy        179 CTTTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTT 230
              | | | || | | |||| |||    |||||  |||| ||||    | | |||
Db         61 CATCTTTAAACAGTCTACACCGAAAACATTTTTGGAAACATCTTTTCCTTTT 112


RESULT 10
ABL33688
ID   ABL33688 standard; DNA; 9646 BP.
XX
AC   ABL33688;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1661.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 1661; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 9646 BP; 2765 A; 150 C; 2136 G; 4595 T; 0 other;

  Query Match             7.9%;  Score 39.8;  DB 24;  Length 9646;
  Best Local Similarity   57.7%;  Pred. No. 0.95;
  Matches   71;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy         72 AGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTTACA 131
              | | ||  || || || |||| || |||          | | | |  |||||||||  |
Db       6728 AAATTTGTTGTTGATATTTTTATTTTTAGGTTTTTTAGTTTTAGTTTATTTTTTTTTGGA 6787

Qy        132 GGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAGAAAT 191
              |  || | || ||  || ||||     | || | | | || ||||  |||| ||| | |
Db       6788 GTGATTTTATTATAATGTAAATTTGATTGTATAGGAAATTTATTTTAATTTTTTAGTAGT 6847

Qy        192 TCT 194
```

```
                | |
Db       6848 TTT 6850


RESULT 11
AAC66096
ID   AAC66096 standard; cDNA; 2360 BP.
XX
AC   AAC66096;
XX
DT   13-FEB-2001  (first entry)
XX
DE   pSS3 cDNA sequence used in cold resistant plant production SEQ ID 6.
XX
KW   Cold resistance; transgenic plant; sucrose decomposition; rice; ss.
XX
OS   Oryza sativa.
XX
PN   JP2000245279-A.
XX
PD   12-SEP-2000.
XX
PF   01-MAR-1999;   99JP-0052102.
XX
PR   01-MAR-1999;   99JP-0052102.
XX
PA   (MITA ) MITSUI CHEM INC.
XX
DR   WPI; 2000-675173/66.
DR   P-PSDB; AAY85666.
XX
PT   Novel method for the reinforcement of cold resistance in a plant
PT   comprising introducing a vector encoding an enzyme that decomposes
PT   sucrose into the plant -
XX
PS   Example 1; Page 15-18; 22pp; Japanese.
XX
CC   This invention relates to a method for the reinforcement of cold
CC   resistance in a plant, comprising introducing an expression vector having
CC   a DNA encoding an enzyme for decomposing sucrose connected downstream to
CC   a promoter expressible in the plant, and expressing the enzyme in the
CC   plant body. Included in the invention are an expression vector used in
CC   the method; a transformed plant carrying the expression vector; and a
CC   transformed rice carrying the above expression vector. The method is used
CC   for reinforcing cold resistance in a plant. The present sequence
CC   represents cDNA used in the production of cold resistant plants by the
CC   method of the invention.
XX
SQ   Sequence 2360 BP; 595 A; 556 C; 599 G; 610 T; 0 other;

  Query Match             7.9%;  Score 39.6;  DB 21;  Length 2360;
  Best Local Similarity   54.0%;  Pred. No. 0.59;
  Matches   81;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy        173 TTTTTCCTTTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTT 232
              |||  ||  | | |    ||   |  | |||| | ||||||  |  ||| ||||||||||
Db        742 TTTGCCCAATCCAATGTGTTGGGATACCCTGATACTGGTGGTCAGGTTGTGTACATTTTG 801

Qy        233 TCCCATATCCCTACTTTTCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGC 292
                |||  |||   ||||  |||| ||  | || |    | |||   | ||  ||
Db        802 GACCAAGTCCGCGCTTTGGAGAATGAGATGCTTTTGAGGATCAAGCAGCAAGGCCTTGAT 861

Qy        293 GTCAGTCTTTTCTTCCTCATCTCTTTCAGG 322
              |||  | |    |||||||      ||||
Db        862 ATCACACCTAAGATCCTCATTGTAACCAGG 891


RESULT 12
AAC66090
ID   AAC66090 standard; cDNA; 2494 BP.
XX
AC   AAC66090;
XX
DT   13-FEB-2001  (first entry)
XX
DE   Rice sucrose synthase cDNA sequence.
XX
KW   Cold resistance; transgenic plant; sucrose decomposition; rice;
KW   sucrose synthase; ss.
XX
OS   Oryza sativa.
XX
PN   JP2000245279-A.
XX
PD   12-SEP-2000.
XX
PF   01-MAR-1999;   99JP-0052102.
XX
PR   01-MAR-1999;   99JP-0052102.
XX
PA   (MITA ) MITSUI CHEM INC.
XX
DR   WPI; 2000-675173/66.
DR   P-PSDB; AAY85664.
XX
PT   Novel method for the reinforcement of cold resistance in a plant
PT   comprising introducing a vector encoding an enzyme that decomposes
PT   sucrose into the plant -
XX
PS   Claim 1; Page 8-11; 22pp; Japanese.
XX
CC   This invention relates to a method for the reinforcement of cold
CC   resistance in a plant, comprising introducing an expression vector having
CC   a DNA encoding an enzyme for decomposing sucrose connected downstream to
CC   a promoter expressible in the plant, and expressing the enzyme in the
CC   plant body. Included in the invention are an expression vector used in
CC   the method; a transformed plant carrying the expression vector; and a
CC   transformed rice carrying the above expression vector. The method is used
CC   for reinforcing cold resistance in a plant. The present sequence
CC   represents cDNA encoding sucrose synthase used in the method.
XX
SQ   Sequence 2494 BP; 622 A; 599 C; 628 G; 645 T; 0 other;

  Query Match             7.9%;  Score 39.6;  DB 21;  Length 2494;
  Best Local Similarity   54.0%;  Pred. No. 0.61;
  Matches   81;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy        173 TTTTTCCTTTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTT 232
              |||  ||  | | |    ||   |  | |||| | ||||||  |  ||| ||||||||||
Db        876 TTTGCCCAATCCAATGTGTTGGGATACCCTGATACTGGTGGTCAGGTTGTGTACATTTTG 935

Qy        233 TCCCATATCCCTACTTTTCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGC 292
                |||  |||   ||||  |||| ||  | || |    | |||   | ||  ||
Db        936 GACCAAGTCCGCGCTTTGGAGAATGAGATGCTTTTGAGGATCAAGCAGCAAGGCCTTGAT 995

Qy        293 GTCAGTCTTTTCTTCCTCATCTCTTTCAGG 322
              |||  | |    |||||||      ||||
Db        996 ATCACACCTAAGATCCTCATTGTAACCAGG 1025


RESULT 13
AAK69902/c
ID   AAK69902 standard; DNA; 648 BP.
XX
AC   AAK69902;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24714.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
```

XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX

```
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 24714; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 648 BP; 233 A; 96 C; 122 G; 197 T; 0 other;

  Query Match             7.7%;  Score 38.6;  DB 22;  Length 648;
  Best Local Similarity   53.3%;  Pred. No. 0.66;
  Matches  105;  Conservative    0;  Mismatches   89;  Indels    3;  Gaps    1;

Qy         86 TACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTTACAGGCATATGAATAAA 145
              ||  |||  |  ||| || |  |||||||||| || ||||| |||    ||| |   |
Db        282 TATATTTGATGTCAGCAACCTTCTTCCTTGATTTTATTTTTCTACTTCAATAAGCCCACT 223

Qy        146 AACTATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAGAAATTCTAAACCTCTGAC 205
              |  |||| ||    | ||  |  |||| || |  |||| |  |    | |   ||
Db        222 ACTTATACTTGTGTGTATGTTTTCTTTTTCCTCTTACTCTATATTTAAATACCACATGCT 163

Qy        206 ATTG---GTGGAGACATTGAGTACATTTTTTCCCATATCCCTACTTTTCAGAAGGATTTT 262
                  |   | || | || |   | |||||| |   | |      | ||||||||  ||
Db        162 TGAGAAAGAAGATAAATAGGTAATATTTTTGCATTTGTAGAAGTTCTTCAGAAGCATACA 103

Qy        263 CTCTGCTCGTTCACTTA 279
                || |||| | ||  ||
Db        102 GACTCCTCGATGACGTA 86


RESULT 14
AAV54123
ID   AAV54123 standard; DNA; 2857 BP.
XX
AC   AAV54123;
XX
DT   17-DEC-1998  (first entry)
XX
DE   Rabbit membrane protein BA0306 coding sequence.
XX
KW   Membrane protein; BA0306; BA2303; arteriosclerosis; coronary restenosis;
KW   therapy; rabbit; ds.
XX
OS   Oryctolagus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             415..2199
FT                   /*tag=  a
FT                   /transl_except= (pos: 745..747, aa: Xaa)
FT                   /note= "Xaa= unspecified amino acid"
XX
PN   WO9838305-A1.
XX
PD   03-SEP-1998.
XX
PF   27-FEB-1998;   98WO-JP00835.
XX
PR   25-FEB-1998;   98JP-0062263.
PR   28-FEB-1997;   97JP-0062259.
XX
PA   (NISB ) JAPAN TOBACCO INC.
XX
PI   Nakamura Y,  Tanaka T,  Tsukada S;
XX
DR   WPI; 1998-481206/41.
DR   P-PSDB; AAW74579.
XX
PT   Membrane protein(s) BA0306 and BA2303 - useful for, e.g. treatment
PT   and prevention of arteriosclerosis and restenosis
XX
PS   Example 4; Page 80-85; 141pp; Japanese.
XX
CC   This sequence represents the DNA encoding the rabbit
CC   membrane protein BA0306. This sequence was used to isolate the human
CC   membrane proteins BA0306 and BA2303 of the invention. The two membrane
CC   proteins are specifically expressed in mammals during arteriosclerosis
CC   and coronary restenosis. The membrane proteins, fragments of them, and
CC   antibodies against them are useful in the treatment and prevention of
CC   arteriosclerosis and restenosis. Transgenic mice expressing the
CC   extracellular region of the membrane proteins are useful as models for
CC   studying these disorders.
XX
SQ   Sequence 2857 BP; 787 A; 553 C; 588 G; 927 T; 2 other;

  Query Match             7.6%;  Score 38.2;  DB 19;  Length 2857;
  Best Local Similarity   52.9%;  Pred. No. 1.6;
  Matches   82;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy         37 GAAAGGCATTGGGAGTCAGCTCCTAAGTAAGCTCCAGAATTCCTGCTGGTACTTTTCCTT 96
              ||||  | |       |   | | ||||  |||  | | | | | |    ||    || ||
Db       2451 GAAATGGACACAATGATGGATTCTAATGAAGACCAAAAGTACTTCTGTGTTTCCTTTCTG 2510

Qy         97 CCAGGAAGCAACTTCCTTGATATTTTTTTTTTACAGGCATATGAATAAAAACTATATTTT 156
              ||| ||||| || | |||  | | | | ||||||  |  |||  |  || | | | ||
Db       2511 TCAGAAAGCATCTCCATTGTAAATATGTATTTACATGTTTATTACAAAGATCCAAATGAA 2570

Qy        157 GCAGCATTGTACACTTTTTTTCCTTTTCTAGAAAT 191
                 |   ||   |  |||||| | |   ||| | ||
Db       2571 AAATTTTTAGTCCATTTTTTGCATAGCCTAAAGAT 2605


RESULT 15
AAS45388
ID   AAS45388 standard; DNA; 15732 BP.
XX
AC   AAS45388;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Chemically pretreated genomic DNA associated with cell cycle #47.
XX
KW   Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW   human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW   graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW   arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW   immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW   PCR primer.
XX
OS   Homo sapiens.
XX
PN   WO200168911-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02945.
XX
PR   15-MAR-2000; 2000DE-1013847.
```

PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602751/68.
XX
PT Designing primers and probes for analysing diseases associated with
PT cytosine methylation state e.g. arthritis, cancer, aging,
PT arteriosclerosis comprising fragments of chemically modified genes
PT associated with cell cycle -
XX
PS Claim 1; SEQ ID No 93; 28pp; English.
XX
CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all CpG dinucleotides in a sequence and therefore for analysing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers.
XX
SQ Sequence 15732 BP; 4638 A; 70 C; 2672 G; 8352 T; 0 other;

Query Match 7.6%; Score 38.2; DB 22; Length 15732;
Best Local Similarity 58.3%; Pred. No. 3.3;
Matches 67; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

```
Qy        119 TTTTTTTTTTACAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTC 178
              |||| ||||||      |||||  ||||| |||||||||| |  ||| || || |||  |
Db      10649 TTTTATTTTTATTTTATTATGATAAAAAAATATATTTTGTATGATTTTAAACGTTTAATA 10708

Qy        179 CTTTTCTAGAAATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTT 233
              || | | | | |    | | | ||| | | |  ||| |   || ||||
Db      10709 TTTATTAATATTTGTTTGGTATTTTAATTTGTGGTATATGTTGGGATAATGTTTT 10763
```

Search completed: November 7, 2003, 06:15:41
Job time : 157.728 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 6, 2003, 21:13:51 ; Search time 155.728 Seconds
(without alignments)
8684.478 Million cell updates/sec

Title: US-09-939-209A-3_COPY_15000_15500
Perfect score: 501
Sequence: 1 tggcagagaactctctggat.....tcaacctgatggagaaggat 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	501	100.0	20300	24	ABK47337	Genomic nucleotide
	2	113	22.6	800	25	ACA56493	Human signalling p
	3	113	22.6	1238	22	AAL26766	Human breast cance
	4	113	22.6	2934	24	ABK47336	cDNA encoding huma
	5	113	22.6	3165	25	ABX63113	Human cDNA #113 di
	6	88.2	17.6	201	18	AAT50916	Rat brain rgss-1 g
	7	60	12.0	60	24	ABN42053	Human spliced tran
	8	52.6	10.5	10427	24	ABK69943	Human secreted pro
	9	49	9.8	378	24	ABL99976	Rat disease associ
	10	49	9.8	543	24	ABL99975	Rat disease associ
	11	49	9.8	618	22	ABA09731	Human bone marrow
	12	49	9.8	817	24	ABL99979	Rat disease associ
	13	48	9.6	540	24	ABA92505	Human regulator of
	14	48	9.6	606	21	AAA39678	Cytokine response
	15	48	9.6	746	19	AAV34780	Human RATH1.1 DNA.
	16	48	9.6	2383	17	AAT11418	p53 response prote
	17	48	9.6	2406	18	AAT43376	Human cytokine res
	18	48	9.6	2406	21	AAA39660	Human CR1 cDNA. H
	19	48	9.6	2434	24	AAS94857	Human DNA sequence
	20	47.4	9.5	745	19	AAV38084	Human regulator of
	21	47.4	9.5	1691	21	AAZ36910	cDNA encoding a re
	22	47.4	9.5	1923	20	AAX51745	DNA encoding a hum
	23	47.4	9.5	1923	24	ABQ92597	Human secreted pro
	24	47.4	9.5	2075	25	ABX74396	Human cDNA sequenc
	25	46.4	9.3	597	22	AAF57422	Human RGS8LIKE pol
	26	46.4	9.3	939	24	ABZ11471	Human polynucleoti
	27	45.8	9.1	955	22	AAI97781	Human neuroblastom
	28	44.8	8.9	2272	19	AAV34779	Mouse RATH1.1 DNA.
	29	43.8	8.7	1396	23	ABV21629	Human prostate exp
	30	43.8	8.7	1396	23	ABV27449	Human prostate exp
	31	43.8	8.7	7345	24	ABL62354	Colon adenocarcino
	32	43.2	8.6	1978	21	AAC79846	Human secreted pro
	33	42.2	8.4	411	25	ABX35169	Bovine EST associa
	34	41.8	8.3	1345	24	ABT10881	Human breast cance
	35	41.8	8.3	1345	24	ABK83834	Human cDNA differe
	36	41.8	8.3	1345	25	ACC46750	Human COPD related
	37	41.8	8.3	1345	25	ACA56680	Signalling pathway
	38	41.8	8.3	1364	21	AAF16132	Human prostate can
	39	41.6	8.3	2638	22	AAH02909	Human shear stress
	40	41.6	8.3	2638	24	ABL62703	Colon adenocarcino
	41	41.6	8.3	2638	24	ABL66645	Lung cancer relate
	42	41.6	8.3	2638	25	ACA56492	Human signalling p
	43	41.6	8.3	2874	25	ABX72245	Human NOVX polynuc
c	44	41	8.2	342	22	AAK54581	Human haematologic
c	45	41	8.2	342	22	AAK54733	Human haematologic

ALIGNMENTS

RESULT 1
ABK47337
ID ABK47337 standard; DNA; 20300 BP.
XX
AC ABK47337;
XX
DT 18-JUN-2002 (first entry)
XX
DE Genomic nucleotide sequence encoding human RGS-4 protein.
XX
KW RGS-4; schizophrenia; human; regulator of G protein signalling 4;
KW neuroleptic; polymorphism; schizotypy; schizoid; chromosome 1q21-22;
KW gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace (4121,T)
FT /*tag= a
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4123,A)
FT /*tag= b
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4368,C)
FT /*tag= c
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4621,C)
FT /*tag= d
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4790,T)
FT /*tag= e

```
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (4816,T)
FT                   /*tag= f
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (4970,T)
FT                   /*tag= g
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace 5055,G)
FT                   /*tag= h
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (5295,A)
FT                   /*tag= i
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (5695,A)
FT                   /*tag= j
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (7375,T)
FT                   /*tag= k
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (7759,A)
FT                   /*tag= l
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (8596,A)
FT                   /*tag= m
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (9602..9610,CA)
FT                   /*tag= n
FT                   /note= "Deletion polymorphism"
FT   variation       replace (9892,A)
FT                   /*tag= o
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (9963,A)
FT                   /*tag= p
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (10132,A)
FT                   /*tag= q
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11056,C)
FT                   /*tag= r
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11091,T)
FT                   /*tag= s
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11106,A)
FT                   /*tag= t
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11774,T)
FT                   /*tag= u
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (12143,A)
FT                   /*tag= v
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (12145,T)
FT                   /*tag= w
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (14367,G)
FT                   /*tag= x
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (17027..17029,GA)
FT                   /*tag= y
FT                   /note= "Deletion polymorphism"
FT   variation       replace (17630,T)
FT                   /*tag= Z
FT                   /standard_name= "Single-nucleotide polymorphism"
XX
PN   WO200216653-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26622.
XX
PR   24-AUG-2000; 2000US-228021P.
XX
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Levitt PR, Mirnics K, Kodavali VC, Nimgaonkar VL;
XX
DR   WPI; 2002-292070/33.
XX
PT   Diagnosing, assessing susceptibility and treating schizophrenia,
PT   involves observing regulator of G-protein signalling 4, RGS4 levels in a
PT   subject -
XX
PS   Claim 1; Page 20-33; 112pp; English.
XX
CC   This invention relates to a novel method for diagnosing schizophrenia
CC   or determining susceptibility to schizophrenia in a human. The method
CC   comprises obtaining from a patient a DNA sample and detecting variations
CC   in the regulator of G-protein signalling 4 (RGS4) gene. Alternatively,
CC   the method involves measuring RGS4 mRNA or protein levels in a tissue
CC   sample from the patient and determining if there is a reduced level.
CC   The method of the invention is useful for diagnosing and determining
CC   susceptibility to schizophrenia. The invention also comprises a method
CC   that is useful for treating schizophrenia which includes a prophylactic
CC   treatment. The method of genotyping polymorphic variants in the RGS-4
CC   gene is applied to diagnosing pathologies of the schizophrenic spectrum,
CC   such as in particular schizotypy, schizoid individuals, etc. This
CC   method offers the possibility of diagnosing schizophrenia by a
CC   biological test and no longer exclusively by clinical evaluations.
CC   The present sequence represents the genomic DNA encoding the human
CC   regulator of G-protein signalling 4 (RGS4) protein used in the method of
CC   the invention. The gene for the RGS4 protein is located on human
CC   chromosome 1q21-22.
XX
SQ   Sequence 20300 BP; 6157 A; 4102 C; 3775 G; 6266 T; 0 other;

  Query Match             100.0%;  Score 501;  DB 24;  Length 20300;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-146;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGGCAGAGAACTCTCTGGATCTTAGTGAAGGTTCCTAGAATAGTGGAGCTGACTATCATA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15000 TGGCAGAGAACTCTCTGGATCTTAGTGAAGGTTCCTAGAATAGTGGAGCTGACTATCATA 15059

Qy         61 ATCTTGACAACCCCAAATAAATCAGTTTTTTAAAAAATCTCTTTTATCCATGTGGCTTAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15060 ATCTTGACAACCCCAAATAAATCAGTTTTTTAAAAAATCTCTTTTATCCATGTGGCTTAC 15119

Qy        121 CATAACCTCCCTGCATGAATTTTTCTGATGAATCTCCCCAATTTGTTAGACAGAACAGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15120 CATAACCTCCCTGCATGAATTTTTCTGATGAATCTCCCCAATTTGTTAGACAGAACAGAA 15179

Qy        181 GATCTTGCCCTGCTCTCTCTAAAGCAGAAAGGTTCATTCTGAACCTTTCATACTCTCTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15180 GATCTTGCCCTGCTCTCTCTAAAGCAGAAAGGTTCATTCTGAACCTTTCATACTCTCTCA 15239

Qy        241 CATGTGCCAAGGAGGACCCCAATGTCACTTTTGTTTTTTGCTTCTGAAATACAGAGGGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15240 CATGTGCCAAGGAGGACCCCAATGTCACTTTTGTTTTTTGCTTCTGAAATACAGAGGGTG 15299

Qy        301 CACTGCCACTTACAAGTCACTACAAAGCATACAGGCTTGCATCCTCAACAGGGATATAGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15300 CACTGCCACTTACAAGTCACTACAAAGCATACAGGCTTGCATCCTCAACAGGGATATAGG 15359

Qy        361 TCTAATGAAGCCTTGGCCTTTGCCCCTCAGGTGAACCTGGATTCTTGCACCAGGGAAGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15360 TCTAATGAAGCCTTGGCCTTTGCCCCTCAGGTGAACCTGGATTCTTGCACCAGGGAAGAG 15419

Qy        421 ACAAGCCGGAACATGCTAGAGCCTACAATAACCTGCTTTGATGAGGCCCAGAAGAAGATT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15420 ACAAGCCGGAACATGCTAGAGCCTACAATAACCTGCTTTGATGAGGCCCAGAAGAAGATT 15479

Qy        481 TTCAACCTGATGGAGAAGGAT 501
              |||||||||||||||||||||
Db      15480 TTCAACCTGATGGAGAAGGAT 15500
```

```
RESULT 2
ACA56493
ID   ACA56493 standard; cDNA; 800 BP.
XX
AC   ACA56493;
XX
DT   06-JUN-2003  (first entry)
XX
DE   Human signalling pathway polynucleotide probe SEQ ID NO 1091.
XX
KW   Human; probe; ss; array element; Parkinson's disease;
KW   signalling pathway population; cancer; adenocarcinoma; leukaemia;
KW   immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.
XX
OS   Homo sapiens.
XX
PN   US6500938-B1.
XX
PD   31-DEC-2002.
XX
PF   30-JAN-1998;   98US-0016434.
XX
PR   30-JAN-1998;   98US-0016434.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Au-Young J,  Seilhamer JJ;
XX
DR   WPI; 2003-352189/33.
XX
PT   Combination of polynucleotide probes, useful as array elements in a
PT   microarray for monitoring the expression of a number of target
PT   polynucleotides  -
XX
PS   Claim 1; SEQ ID NO 1091; 65pp; English.
XX
CC   The invention relates to a combination which, comprises a number of
CC   polynucleotide probes comprising a sequence selected from one of the 1490
CC   sequences mentioned in the specification. The combination is useful as an
CC   array element in a microarray for monitoring the expression of a number
CC   of target polynucleotides. The microarray is particularly useful in the
CC   diagnosis and treatment of cancer and immunopathology and neuropathology.
CC   The microarray is useful in diagnostics and treatment regimens, drug
CC   discovery and development, toxicological and carcinogenicity studies,
CC   forensics and pharmacogenomics. The microarray is also useful for
CC   monitoring progression of diseases and for developing sophisticated
CC   profiles for the effects of currently available therapeutic drugs. The
CC   combination is also useful for purifying a subpopulation of mRNAs, cDNAs
CC   and genomic fragments and in research and diagnostic applications. The
CC   array can detect changes in expression in a large number of genes coding
CC   for different signaling pathway populations which can be used to diagnose
CC   various diseases including cancer e.g. adenocarcinoma and leukaemia,
CC   immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease
CC   and Parkinson's disease. The present sequence represents a polynucleotide
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=06500938B1.
XX
SQ   Sequence 800 BP; 241 A; 181 C; 195 G; 183 T; 0 other;

  Query Match             22.6%;  Score 113;  DB 25;  Length 800;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-25;
  Matches  113;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        474 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 533

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 501
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        534 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 586
```

```
RESULT 3
AAL26766
ID   AAL26766 standard; cDNA; 1238 BP.
XX
AC   AAL26766;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 19223.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer  -
XX
PS   Claim 1; Page 3683-3684; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 1238 BP; 386 A; 265 C; 307 G; 274 T; 6 other;

  Query Match             22.6%;  Score 113;  DB 22;  Length 1238;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-25;
  Matches  113;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        674 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 733

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 501
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        734 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 786


RESULT 4
ABK47336
ID   ABK47336 standard; DNA; 2934 BP.
XX
AC   ABK47336;
XX
DT   18-JUN-2002  (first entry)
XX
```

```
DE   cDNA encoding human RGS-4 protein.
XX
KW   RGS-4; schizophrenia; human; regulator of G protein signalling 4;
KW   neuroleptic; polymorphism; schizotypy; schizoid; chromosome 1q21-22;
KW   gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             77..694
FT                   /*tag=  a
FT                   /product= "RGS-4 protein"
XX
PN   WO200216653-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26622.
XX
PR   24-AUG-2000; 2000US-228021P.
XX
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Levitt PR,  Mirnics K,  Kodavali VC,  Nimgaonkar VL;
XX
DR   WPI; 2002-292070/33.
DR   P-PSDB; AAU78976.
XX
PT   Diagnosing, assessing susceptibility and treating schizophrenia,
PT   involves observing regulator of G-protein signalling 4, RGS4 levels in a
PT   subject  -
XX
PS   Disclosure; Page 18-20; 112pp; English.
XX
CC   This invention relates to a novel method for diagnosing schizophrenia
CC   or determining susceptibility to schizophrenia in a human. The method
CC   comprises obtaining from a patient a DNA sample and detecting variations
CC   in the regulator of G-protein signalling 4 (RGS4) gene. Alternatively,
CC   the method involves measuring RGS4 mRNA or protein levels in a tissue
CC   sample from the patient and determining if there is a reduced level.
CC   The method of the invention is useful for diagnosing and determining
CC   susceptibility to schizophrenia. The invention also comprises a method
CC   that is useful for treating schizophrenia which includes a prophylactic
CC   treatment. The method of genotyping polymorphic variants in the RGS-4
CC   gene is applied to diagnosing pathologies of the schizophrenic spectrum,
CC   such as in particular schizotypy, schizoid individuals, etc. This
CC   method offers the possibility of diagnosing schizophrenia by a
CC   biological test and no longer exclusively by clinical evaluations.
CC   The present sequence represents the cDNA encoding the human regulator of
CC   G-protein signalling 4 (RGS4) protein used in the method of the
CC   invention. The gene for the RGS4 protein is located on human chromosome
CC   1q21-22.
XX
SQ   Sequence 2934 BP; 818 A; 594 C; 599 G; 923 T; 0 other;

  Query Match             22.6%;  Score 113;  DB 24;  Length 2934;
  Best Local Similarity   100.0%;  Pred. No. 8e-25;
  Matches  113;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        453 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 512

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 501
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        513 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 565


RESULT 5
ABX63113
ID   ABX63113 standard; cDNA; 3165 BP.
XX
AC   ABX63113;
XX
DT   25-FEB-2003  (first entry)
XX
DE   Human cDNA #113 differentially expressed in activated vascular tissue.
XX
KW   Human; gene; ss; vascular tissue; cytostatic; atherosclerosis;
KW   cardiant; hypotensive; antidiabetic; gynaecological; vasotropic;
KW   cerebroprotective; gene therapy; vascular disease; cancer; coronary;
KW   artery disease; hypertension; diabetes; pre-eclampsia; restenosis;
KW   ischaemia-reperfusion injury; stroke;
XX
OS   Homo sapiens.
XX
PN   US2002137081-A1.
XX
PD   26-SEP-2002.
XX
PF   08-JAN-2002; 2002US-0044090.
XX
PR   28-JUL-2000; 2000US-222469P.
PR   08-JAN-2001; 2001US-260483P.
XX
PA   (BAND/) BANDMAN O.
XX
PI   Bandman O;
XX
DR   WPI; 2003-110597/10.
XX
PT   Combination for diagnosing, staging, treating, or monitoring the
PT   progression of treatment of a vascular disease, e.g. atherosclerosis,
PT   comprises several cDNAs that are differentially expressed in activated
PT   vascular tissue -
XX
PS   Claim 1; Page -; 18pp; English.
XX
CC   This invention relates to a combination comprising several cDNAs that
CC   are differentially expressed in activated vascular tissue. The invention
CC   also discloses a high throughput method for detecting differentially
CC   expressed cDNAs in a sample. The cDNAs of the invention may have
CC   antiarteriosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;
CC   gynaecological; vasotropic and cerebroprotective activities and may be
CC   used in gene therapy. The cDNAs of the invention may be used in a
CC   high-throughput methods for detecting differential expression of one or
CC   more cDNAs in a sample, or screening several molecules or compounds to
CC   identify a molecule or compound that specifically binds a cDNA of the
CC   invention. A protein encoded by the cDNA may be used to screen several
CC   molecules or compounds to identify a ligand that specifically binds to
CC   the protein, or to produce or purify an antibody to the protein that can
CC   be used to detect a protein in a sample or purify a natural or
CC   recombinant protein from a sample. The nucleotides may be useful for
CC   diagnosing, staging, treating, or monitoring the progression of
CC   treatment of a vascular disease, e.g. atherosclerosis, cancer, coronary
CC   artery disease, hypertension, diabetes, pre-eclampsia, ischaemia-
CC   reperfusion injury, restenosis, or stroke. The cDNAs can also be used
CC   for large-scale genetic or gene expression analysis of several new
CC   nucleic acid molecules. Antibodies to the proteins encoded by the
CC   cDNAs are useful for diagnosing pre-pathologic disorders, and chronic
CC   or acute diseases associated with abnormalities in the expression,
CC   amount or distribution of the protein. The present sequence
CC   represents a cDNA of the invention that is differentially expressed in
CC   activated vascular tissue.
CC   Note: The sequence data for this patent did not form part of the
CC   specification, but was obtained in electronic format directly from USPTO
CC   at http.seqdata.uspto.gov/sequence.html?DocID=20020137081.
XX
SQ   Sequence 3165 BP; 889 A; 636 C; 672 G; 967 T; 1 other;

  Query Match             22.6%;  Score 113;  DB 25;  Length 3165;
  Best Local Similarity   100.0%;  Pred. No. 8.3e-25;
  Matches  113;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        679 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 738

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 501
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        739 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 791


RESULT 6
AAT50916
ID   AAT50916 standard; DNA; 201 BP.
XX
AC   AAT50916;
XX
DT   31-MAR-1997  (first entry)
XX
DE   Rat brain rgss-1 gene fragment.
XX
KW   Regulator G-protein signalling similarity; rgss; RGS;
KW   signal transduction; transgenic animal; diagnosis; therapy;
KW   diabetes; hyperplasia; psychiatric disorder;
KW   cardiovascular disease; McCune-Albright syndrome;
KW   Albright hereditary osteopathy; ss.
XX
OS   Rattus sp.
XX
PN   WO9638462-A1.
XX
PD   05-DEC-1996.
XX
PF   31-MAY-1996;   96WO-US08295.
XX
PR   12-JAN-1996;   96US-0588258.
PR   02-JUN-1995;   95US-0460505.
XX
PA   (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
PI   Horvitz HR,  Koelle M;
XX
DR   WPI; 1997-034298/03.
DR   P-PSDB; AAW10169.
XX
PT   New isolated regulator of G-protein signalling genes - used to
PT   develop prods. for the diagnosis and treatment of G-protein related
PT   diseases and disorders e.g. diabetes, cardiovascular disease, etc
XX
PS   Example B; Page 53; 96pp; English.
XX
CC   Gene fragments (AAT50916-24), designated rgss-1 to rgss-9, were
CC   isolated from rat brain cDNA using degenerate primers (AAT50912-15)
CC   based on the conserved region of nematode EGL-10 protein (see
CC   also AAW10167 and AAW10178).  EGL-10 is a member of the new RGS
CC   (Regulators of G-protein Signalling) family involved in the control
CC   of G-protein mediated effects.  The rat gene fragments, designated
CC   rgss for regulator G-protein signalling similarity, encode proteins
CC   (AAW10169-77) contg. RGS conserved regions.  RGS genes (see also
CC   AAT50910-11) can be used to detect related genes, to produce RGS
CC   polypeptides in transformed host cells, and in the diagnosis and
CC   treatment of G-protein related disorders.
XX
SQ   Sequence 201 BP; 70 A; 44 C; 46 G; 41 T; 0 other;

  Query Match            17.6%;  Score 88.2;  DB 18;  Length 201;
  Best Local Similarity   92.1%;  Pred. No. 1.4e-17;
  Matches   93;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              |||||||||||||||||||||||| || |||||||||||||||||| ||||||| || |
Db        101 AGGTGAACCTGGATTCTTGCACCAGAGAGGAGACAAGCCGGAACATGTTAGAGCCCACGA 160

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTG 489
              ||||||| |||||||| |||| ||||||||||||||||||
Db        161 TAACCTGTTTTGATGAAGCCCGGAAGAAGATTTTCAACCTG 201


RESULT 7
ABN42053
ID   ABN42053 standard; DNA; 60 BP.
XX
AC   ABN42053;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human spliced transcript detection oligonucleotide SEQ ID NO:14801.
XX
KW   Human; mouse; rat; splice transcript; detection; RNA transcript;
KW   splice variant; transcriptome; oligonucleotide library; ss.
XX
OS   Homo sapiens.
XX
PN   WO200210449-A2.
XX
PD   07-FEB-2002.
XX
PF   20-JUL-2001; 2001WO-IB01903.
XX
PR   28-JUL-2000; 2000US-221607P.
PR   02-MAY-2001; 2001US-287724P.
XX
PA   (COMP-) COMPUGEN INC.
XX
PI   Shoshan A,  Wasserman A,  Mintz E,  Mintz L,  Faigler S;
XX
DR   WPI; 2002-257383/30.
XX
PT   New oligonucleotide libraries comprising oligonucleotides which
PT   selectively hybridize to mRNAs transcribed from a transcription unit of
PT   a genome, useful for detecting tissue-, pathology-, and
PT   developmental-specific genes  -
XX
PS   Example 1; SEQ ID 14801; 47pp; English.
XX
CC   The present invention describes oligonucleotide libraries for detecting
CC   messenger RNAs that populate a (sub-)transcriptome, where the
CC   (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC   transcription units that populate a genome. The library comprises
CC   several oligonucleotides, each capable of hybridising selectively to a
CC   set of messenger RNAs transcribed from a given transcription unit of
CC   the genome, which encodes one or more messenger RNA splice variants.
CC   The oligonucleotide libraries are useful for detecting mRNAs from a
CC   biological sample, in expression profiling studies, in qualitatively or
CC   quantitatively characterising the corresponding transcriptome, and in
CC   detecting RNA transcripts and splice variants of human or animal
CC   transcriptomes. The libraries may also be used as specialised mini
CC   libraries to detect transcripts of a sub-transcriptome under a
CC   particular biological or pathological state, and so allowing the
CC   detection of tissue- and pathology-specific genes such as those genes
CC   only expressed in specific tissue under a specific pathological
CC   condition; to detect developmental specific genes; and to detect RNA
CC   transcripts and splice variants of a transcriptome of a patient suffering
CC   from a particular disorder. ABN27253 to ABN59589 represent
CC   oligonucleotide sequences from rats, humans and mice, which are used in
CC   the exemplification of the present invention.
CC   N.B. The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 60 BP; 21 A; 14 C; 16 G; 9 T; 0 other;

  Query Match            12.0%;  Score 60;  DB 24;  Length 60;
  Best Local Similarity  100.0%;  Pred. No. 6.2e-09;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        418 GAGACAAGCCGGAACATGCTAGAGCCTACAATAACCTGCTTTGATGAGGCCCAGAAGAAG 477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAGACAAGCCGGAACATGCTAGAGCCTACAATAACCTGCTTTGATGAGGCCCAGAAGAAG 60
```

```
RESULT 8
ABK69943
ID   ABK69943 standard; DNA; 10427 BP.
XX
AC   ABK69943;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human secreted protein gene 69 genomic DNA fragment #2.
XX
KW   Human; ds; secreted protein; gene therapy; immunosuppressive;
KW   antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW   vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW   virucide; fungicide; opthalmalogical; autoimmune disease; neoplasm;
KW   rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW   cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW   angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW   ocular disorder; corneal infection; wound healing; skin aging;
KW   epithelial cell proliferation; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200226931-A2.
XX
PD   04-APR-2002.
XX
PF   24-SEP-2001; 2001WO-US29871.
XX
PR   25-SEP-2000; 2000US-234925P.
PR   12-JAN-2001; 2001WO-US00911.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis G,  Duan DR,  Rosen CA,  Moore PA,  Shi Y;
PI   Lafleur DW,  Olsen H,  Brewer LA,  Florence KA,  Young PE,  Soppet DR;
PI   Endress GA,  Mucenski M,  Ebner R;
XX
DR   WPI; 2002-362489/39.
XX
PT   Novel 71 isolated secreted polypeptides and polynucleotides encoding
PT   the polypeptides, useful for treating Huntington's disease, sepsis,
PT   meningitis, thrombocytopaenia, haemolytic anaemia, rheumatoid arthritis,
PT   asthma  -
XX
PS   Example 2; Page 1452-1455; 1478pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (or its
CC   fragment, homologue complement or allelic variant) encoding a human
CC   secreted protein (and its fragment, domain, epitope, variant, secreted
CC   form and species variant). Also included are a recombinant vector
CC   comprising the nucleic acid, a recombinant host cell comprising the
CC   vector, an antibody against the secreted protein, a recombinant host cell
CC   that expresses the secreted protein and a method of identifying a binding
CC   partner of the secreted protein. The nucleic acid and protein are used to
CC   prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC   humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep
CC   for example autoimmune diseases e.g. rheumatoid arthritis,
CC   hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC   cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC   e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC   Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC   ocular disorders e.g. corneal infection. Many other diseases and
CC   disorders are listed in the specification. The polypeptides can also be
CC   used to aid wound healing an epithelial cell proliferation, to prevent
CC   skin aging due to sunburn, to maintain organs before transplantation, for
CC   supporting cell culture of primary tissues, to regenerate tissues and in
CC   chemotaxis. The polypeptides can also be used as a food additive or
CC   preservative to increase or decrease storage capabilities. The present
CC   sequence represents a ds DNA fragment of the gene for a novel human
CC   secreted protein of the invention.
XX
SQ   Sequence 10427 BP; 3337 A; 1995 C; 1829 G; 3266 T; 0 other;

  Query Match             10.5%;  Score 52.6;  DB 24;  Length 10427;
  Best Local Similarity   62.6%;  Pred. No. 1.5e-05;
  Matches   82;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        371 CCTTGGCCTTTGCCCCTCAGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGA 430
              || ||     || ||| ||||||||  | ||    | ||| | |||    ||||    ||
Db       5202 CCCTGTTTCCTGGCCCACAGGTGAATATTGACCACTTCACTAAGGACATCACAATGAAGA 5261

Qy        431 ACATGCTAGAGCCTACAATAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGA 490
              || || | || ||| |  | | | |||||||   |||||||| |  ||      |||||
Db       5262 ACCTGGTGGAACCTTCCCTGAGCAGCTTTGACATGGCCCAGAAAAGAATCCATGCCCTGA 5321

Qy        491 TGGAGAAGGAT 501
              |||| ||||||
Db       5322 TGGAAAAGGAT 5332


RESULT 9
ABL99976
ID   ABL99976 standard; DNA; 378 BP.
XX
AC   ABL99976;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Rat disease associated gene related polynucleotide SEQ ID NO 3.
XX
KW   Rat; cardiant; heart disease; cardiovascular disease; cardiac infarction;
KW   angina; gene therapy; ds.
XX
OS   Rattus sp.
XX
PN   WO200233082-A1.
XX
PD   25-APR-2002.
XX
PF   18-OCT-2001; 2001WO-JP09140.
XX
PR   19-OCT-2000; 2000JP-0319912.
PR   16-NOV-2000; 2000JP-0350183.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Koyama N,  Tanida S,  Watanabe T;
XX
DR   WPI; 2002-394557/42.
XX
PT   Disease-associated gene and encoded RGS5-like protein, applicable in
PT   diagnosis and prevention or treatment of heart diseases e.g.
PT   cardiovascular diseases, cardiac infarction, heart failure and angina,
PT   including gene therapy  -
XX
PS   Example 1; Page 73; 82pp; Japanese.
XX
CC   The invention relates to a protein (ABB83788) or its salt. The protein
CC   and encoded DNA (ABL99975) are applicable in diagnosis and prevention or
CC   treatment of heart diseases e.g. cardiovascular diseases, cardiac
CC   infarction, heart failure and angina, including gene therapy. The present
CC   sequence is that of a polynucleotide, useful in examples of the
CC   invention.
XX
SQ   Sequence 378 BP; 115 A; 101 C; 80 G; 82 T; 0 other;

  Query Match              9.8%;  Score 49;  DB 24;  Length 378;
  Best Local Similarity   64.6%;  Pred. No. 4.2e-05;
  Matches   73;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              |||||||| | ||    | ||| |  ||    || |    |||| || | || ||| |
Db         65 AGGTGAACATTGACCACTTCACTAAAGACATCACCATGAAGAACCTGGTGGAACCTTCCC 124

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 501
```

```
                    | |||||||   |||||||||| | ||| | |  ||||||||||||||||
Db        125 CTCACAGCTTTGACCTGGCCCAGAAAAGGATCTACGCCCTGATGGAGAAGGAT 177
```

RESULT 10
ABL99975
ID ABL99975 standard; cDNA; 543 BP.
XX
AC ABL99975;
XX
DT 14-AUG-2002 (first entry)
XX
DE Rat disease associated gene SEQ ID NO 2.
XX
KW Rat; cardiant; heart disease; cardiovascular disease; cardiac infarction;
KW angina; gene therapy; gene; ss.
XX
OS Rattus sp.
XX
FH Key   Location/Qualifiers
FT CDS   1..543
FT   /*tag= a
FT   /product= "disease associated protein"
XX
PN WO200233082-A1.
XX
PD 25-APR-2002.
XX
PF 18-OCT-2001; 2001WO-JP09140.
XX
PR 19-OCT-2000; 2000JP-0319912.
PR 16-NOV-2000; 2000JP-0350183.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Koyama N, Tanida S, Watanabe T;
XX
DR WPI; 2002-394557/42.
DR P-PSDB; ABB83788.
XX
PT Disease-associated gene and encoded RGS5-like protein, applicable in
PT diagnosis and prevention or treatment of heart diseases e.g.
PT cardiovascular diseases, cardiac infarction, heart failure and angina,
PT including gene therapy -
XX
PS Claim 6; Page 72; 82pp; Japanese.
XX
CC The invention relates to a protein (ABB83788) or its salt. The protein
CC and encoded DNA (ABL99975) are applicable in diagnosis and prevention or
CC treatment of heart diseases e.g. cardiovascular diseases, cardiac
CC infarction, heart failure and angina, including gene therapy.
XX
SQ Sequence 543 BP; 160 A; 137 C; 128 G; 118 T; 0 other;

```
  Query Match             9.8%;  Score 49;  DB 24;  Length 543;
  Best Local Similarity   64.6%;  Pred. No. 5e-05;
  Matches   73;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              ||||||||| | ||    | ||| |  ||     || |    |||| || | || ||| |
Db        383 AGGTGAACATTGACCACTTCACTAAAGACATCACCATGAAGAACCTGGTGGAACCTTCCC 442

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 501
              | |||||||   |||||||| | ||| | |  ||||||||||||||||
Db        443 CTCACAGCTTTGACCTGGCCCAGAAAAGGATCTACGCCCTGATGGAGAAGGAT 495
```

RESULT 11
ABA09731
ID ABA09731 standard; DNA; 618 BP.
XX
AC ABA09731;
XX
DT 15-JAN-2002 (first entry)
XX
DE Human bone marrow expressed oligonucleotide SEQ ID NO: 240.
XX
KW Human; bone marrow; cytostatic; antirheumatic; antiarthritic; vulnerary;
KW antiinflammatory; antibacterial; immunosuppressive; vasotropic; cancer;
KW antiparkinsonian; neuroprotective; nootropic; haemostatic; osteopathic;
KW antiulcer; fungicide; antidiabetic; antiasthmatic; antiallergic;
KW immunostimulant; analgesic; cerebroprotective; antianaemic; infection;
KW nervous system disorder; autoimmune disorder; inflammation; allergy; ds.
XX
OS Homo sapiens.
XX
PN WO200174836-A1.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US10472.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PR 23-AUG-2000; 2000US-0649267.
PR 30-NOV-2000; 2000US-250583P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Ford JE, Boyle BJ;
XX
DR WPI; 2001-626375/72.
XX
PT New bone marrow-expressed nucleic acids and polypeptides, useful for
PT diagnosis, treatment of inflammatory, autoimmune, neurological, cancer
PT and increasing hematopoiesis, stem cell survival and bone growth and
PT remodeling -
XX
PS Claim 1; Page 299; 380pp; English.
XX
CC The present invention relates to bone marrow expressed polynucleotides
CC and proteins. These sequences can be used in the treatment of
CC inflammatory conditions (eg arthritis, Crohn's disease), cancer, central
CC and peripheral nervous system diseases and neuropathies, such as
CC Alzheimer's, Parkinson's and Huntington's diseases, spinal cord
CC disorders, head trauma, cerebrovascular diseases, myeloid and lymphoid
CC cell disorders, platelet disorders, stem cell disorders, bone
CC degenerative disorders, autoimmune disorders, for example multiple
CC sclerosis, diabetes and arthritis, viral and bacterial infections,
CC allergies and blood coagulation disorders. The present sequence is a DNA
CC of the invention.
XX
SQ Sequence 618 BP; 186 A; 143 C; 155 G; 134 T; 0 other;

```
  Query Match             9.8%;  Score 49;  DB 22;  Length 618;
  Best Local Similarity   57.5%;  Pred. No. 5.3e-05;
  Matches   88;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy        348 ACAGGGATATAGGTCTAATGAAGCCTTGGCCTTTGCCCCTCAGGTGAACCTGGATTCTTG 407
              ||  || ||   | | | |    |  | | | |  | ||||| ||| | || |
Db        393 ACCTGGTTACCAGCCAAGAGGTCACACAGTCATGGTGTCACAGGTAAACATTGACTTCCA 452

Qy        408 CACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAATAACCTGCTTTGATGAGGC 467
              ||| | |||| || ||   |||| ||| |||||  |  | || ||||||||  | ||
Db        453 GACCCGAGAAGCCACGAGGAAGAACCTGCAGGAGCCATCCCTGACTTGCTTTGACCAAGC 512

Qy        468 CCAGAAGAAGATTTTCAACCTGATGGAGAAGGA 500
              |||    ||  |  || ||| |||||||| ||
Db        513 CCAAGGAAAAGTACACAGCCTCATGGAGAAAGA 545
```

RESULT 12
ABL99979
ID ABL99979 standard; DNA; 817 BP.
XX
AC ABL99979;

```
XX
DT   14-AUG-2002  (first entry)
XX
DE   Rat disease associated gene related polynucleotide SEQ ID NO 6.
XX
KW   Rat; cardiant; heart disease; cardiovascular disease; cardiac infarction;
KW   angina; gene therapy; ds.
XX
OS   Rattus sp.
XX
PN   WO200233082-A1.
XX
PD   25-APR-2002.
XX
PF   18-OCT-2001; 2001WO-JP09140.
XX
PR   19-OCT-2000; 2000JP-0319912.
PR   16-NOV-2000; 2000JP-0350183.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Koyama N,  Tanida S,  Watanabe T;
XX
DR   WPI; 2002-394557/42.
XX
PT   Disease-associated gene and encoded RGS5-like protein, applicable in
PT   diagnosis and prevention or treatment of heart diseases e.g.
PT   cardiovascular diseases, cardiac infarction, heart failure and angina,
PT   including gene therapy  -
XX
PS   Example 1; Page 74-75; 82pp; Japanese.
XX
CC   The invention relates to a protein (ABB83788) or its salt. The protein
CC   and encoded DNA (ABL99975) are applicable in diagnosis and prevention or
CC   treatment of heart diseases e.g. cardiovascular diseases, cardiac
CC   infarction, heart failure and angina, including gene therapy. The present
CC   sequence is that of a polynucleotide, useful in examples of the
CC   invention.
XX
SQ   Sequence 817 BP; 228 A; 214 C; 192 G; 183 T; 0 other;

  Query Match             9.8%;  Score 49;  DB 24;  Length 817;
  Best Local Similarity   64.6%;  Pred. No. 6.1e-05;
  Matches   73;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              ||||||||  |  ||    |  ||| |  ||    || |    |||| ||  |  || ||| |
Db        504 AGGTGAACATTGACCACTTCACTAAAGACATCACCATGAAGAACCTGGTGGAACCTTCCC 563

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGAT 501
                |  |||||||   |||||||| | ||| | |  ||||||||||||||||
Db        564 CTCACAGCTTTGACCTGGCCCAGAAAAGGATCTACGCCCTGATGGAGAAGGAT 616


RESULT 13
ABA92505
ID   ABA92505 standard; cDNA; 540 BP.
XX
AC   ABA92505;
XX
DT   19-MAR-2002  (first entry)
XX
DE   Human regulator of G protein signalling (RGS8) encoding cDNA.
XX
KW   Human; regulator of G protein signalling; RGS8; cerebroprotective;
KW   vulnerary; tranquilliser; analgesic; anticonvulsant; vasotropic;
KW   vaccine; stroke; head trauma; anxiety; pain; epileptic seizure; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..540
FT                   /*tag=  a
FT                   /partial
FT                   /product= "regulator of G protein signalling (RGS8)"
FT                   /note= "no stop codon given"
XX
PN   WO200185937-A2.
XX
PD   15-NOV-2001.
XX
PF   10-MAY-2001; 2001WO-EP05295.
XX
PR   12-MAY-2000; 2000EP-0110163.
XX
PA   (MERE ) MERCK PATENT GMBH.
XX
PI   Wilm C,  Gassen M;
XX
DR   WPI; 2002-055595/07.
DR   P-PSDB; ABB04999.
XX
PT   Novel regulator of G-protein signaling polypeptide and polynucleotide
PT   for diagnosing, treating stroke, head trauma, anxiety, pain, epileptic
PT   seizures and for identifying modulators of therapeutic use  -
XX
PS   Claim 5; Page 34-35; 35pp; English.
XX
CC   The present sequence encodes a human regulator of G-protein signalling,
CC   designated RGS8. RGS8 has cerebroprotective, vulnerary, tranquilliser,
CC   analgesic, anticonvulsant and vasotropic activities, and can be used in
CC   vaccine production. The RGS8 protein is useful in screening assays to
CC   identify compounds that stimulate or inhibit the function or level of
CC   the protein. RGS8 proteins and polynucleotides are useful as vaccines.
CC   The proteins are useful as immunogens to produce antibodies which are
CC   useful for treating diseases, to isolate or to identify clones
CC   expressing the protein or to purify the proteins by affinity
CC   chromatography. RGS8 proteins are also useful to identify membrane
CC   bound or soluble receptors. The RGS8 polynucleotide is useful for the
CC   recombinant production of RGS8 proteins, as hybridisation probes for
CC   cDNA and genomic DNA, as primers for nucleic acid amplification reaction
CC   to isolate full-length cDNAs and genomic clones encoding RGS8 proteins,
CC   in diagnostic assays by detecting mutations in the associated gene, for
CC   chromosome localisation studies, tissue expression studies and for
CC   producing transgenic animals useful in drug discovery and target
CC   validation. RGS8 proteins and polynucleotides can be used in the
CC   diagnosis and treatment of stroke, head trauma, anxiety, pain, epileptic
CC   seizures and for identifying modulators of therapeutic use.
XX
SQ   Sequence 540 BP; 143 A; 131 C; 147 G; 119 T; 0 other;

  Query Match             9.6%;  Score 48;  DB 24;  Length 540;
  Best Local Similarity   64.3%;  Pred. No. 0.0001;
  Matches   72;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              ||||||||  |  ||  |     |||  |  ||||  ||  ||   ||||  |||  |||||  |
Db        359 AGGTGAACATTGACTTCCAGACCCGAGAAGCCACGAGGAAGAACCTGCAGGAGCCATCCC 418

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGA 500
              |  ||  |||||||||   |  |||||     ||   |    ||  |||  |||||||||  ||
Db        419 TGACTTGCTTTGACCAAGCCCAAGGAAAAGTACACAGCCTCATGGAGAAAGA 470


RESULT 14
AAA39678
ID   AAA39678 standard; cDNA; 606 BP.
XX
AC   AAA39678;
XX
DT   18-SEP-2000  (first entry)
XX
DE   Cytokine response gene-related cDNA sequence (Seq ID 27).
XX
KW   CR2; human; antibody; cytokine response gene; cytostatic; anti-allergic;
KW   immunosuppressive; antimicrobial; therapy; cell proliferation; treatment;
```

```
KW   cell differentiation; cancer; immune disease; rheumatologic disease;
KW   transplant rejection; anti-infective; ss.
XX
OS   Unidentified.
XX
PN   US6057427-A.
XX
PD   02-MAY-2000.
XX
PF   05-JUN-1996;   96US-0652446.
XX
PR   20-NOV-1991;   91US-0796066.
PR   10-AUG-1993;   93US-0104736.
PR   27-OCT-1994;   94US-0330108.
PR   05-JUN-1996;   96WO-US08992.
XX
PA   (DART-) DARTMOUTH COLLEGE.
XX
PI   Beadling C,  Smith KA;
XX
DR   WPI; 2000-338623/29.
XX
PT   Novel antibody or antibody fragment which selectively binds to a
PT   polypeptide encoded by cytokine response gene 2  -
XX
PS   Disclosure; Column 113-116; 66pp; English.
XX
CC   This invention describes a novel isolated antibody or antibody fragment
CC   (I) which selectively binds to a polypeptide encoded by cytokine response
CC   gene 2 (CR2) and modulates CR2 activity. The products of the invention
CC   have cytostatic, anti-allergic, immunosuppressive and antimicrobial
CC   activity. The antibodies are useful as therapeutic agents for regulating
CC   cellular proliferation and differentiation and for treating all kinds of
CC   cancers, immune diseases such as allergic, autoimmune, and rheumatologic
CC   diseases, transplant rejection, and as anti-infectives for fighting
CC   viral, bacterial, parasitic and fungal infections. This sequence is
CC   presented in the Seq ID list but is not described in the specification.
XX
SQ   Sequence 606 BP; 147 A; 172 C; 170 G; 117 T; 0 other;

  Query Match             9.6%;  Score 48;  DB 21;  Length 606;
  Best Local Similarity   64.3%;  Pred. No. 0.00011;
  Matches   72;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              |||| ||| | ||   |   ||| | ||   ||| ||   |||| |||       | |||
Db        386 AGGTCAACATTGACCATGAGACCCGCGAGCTGACGAGGATGAACCTGCAGACTGCCACAG 445

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGA 500
                || |||||||||| ||| |||  |||||     | ||||||||||||||
Db        446 CCACATGCTTTGATGCGGCTCAGGGGAAGACACGTACCCTGATGGAGAAGGA 497


RESULT 15
AAV34780
ID   AAV34780 standard; DNA; 746 BP.
XX
AC   AAV34780;
XX
DT   11-SEP-1998  (first entry)
XX
DE   Human RATH1.1 DNA.
XX
KW   RATH1.1; RGS gene family; regulator of G-protein signalling; human;
KW   regulation; activation; modulation; signal transduction; T Cell;
KW   T helper cell; treatment; autoimmune disorders; arthritis; infection;
KW   graft rejection; asthma; allergy; neoplasm; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             39..647
FT                   /*tag=  a
FT                   /product= RATH1.1
XX
PN   WO9814579-A1.
XX
PD   09-APR-1998.
XX
PF   06-OCT-1997;   97WO-US18259.
XX
PR   06-JUN-1997;   97US-0870815.
PR   04-OCT-1996;   96US-0726228.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Gimeno CJ,  Levinson DA;
XX
DR   WPI; 1998-240085/21.
DR   P-PSDB; AAW59294.
XX
PT   New activated T helper cell specific gene, RATH - used for
PT   developing products for treating e.g. autoimmune disorders,
PT   arthritis, graft rejection, asthma, allergy, infections or neoplasms
XX
PS   Claim 1; Fig 2A-B; 127pp; English.
XX
CC   This sequence encodes a novel member of the human regulator of G-protein
CC   signalling (RGS) gene family, RATH1.1. The encoded protein participates
CC   in the regulation, control and/or modulation of G-protein mediated
CC   signal transduction, involved in T cell activation, including T-helper
CC   (TH) cell and TH cell subpopulation activation. The protein can be
CC   used to develop products for treating autoimmune disorders, arthritis,
CC   graft rejection, asthma, allergy, infections or neoplasms.
XX
SQ   Sequence 746 BP; 172 A; 217 C; 215 G; 142 T; 0 other;

  Query Match             9.6%;  Score 48;  DB 19;  Length 746;
  Best Local Similarity   64.3%;  Pred. No. 0.00012;
  Matches   72;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy        389 AGGTGAACCTGGATTCTTGCACCAGGGAAGAGACAAGCCGGAACATGCTAGAGCCTACAA 448
              |||| ||| | ||   |   ||| | ||   ||| ||   |||| |||       | |||
Db        424 AGGTCAACATTGACCATGAGACCCGCGAGCTGACGAGGATGAACCTGCAGACTGCCACAG 483

Qy        449 TAACCTGCTTTGATGAGGCCCAGAAGAAGATTTTCAACCTGATGGAGAAGGA 500
                || |||||||||| ||| |||  |||||     | ||||||||||||||
Db        484 CCACATGCTTTGATGCGGCTCAGGGGAAGACACGTACCCTGATGGAGAAGGA 535


Search completed: November  7, 2003, 06:15:44
Job time : 158.728 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: November 6, 2003, 21:13:51 ; Search time 155.728 Seconds
(without alignments)
8684.478 Million cell updates/sec

Title: US-09-939-209A-3_COPY_19800_20300
Perfect score: 501
Sequence: 1 ccacatgattatctcaatag.....atgagtgaactcccattcac 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
✓	1	501	100.0	20300	24	ABK47337	Genomic nucleotide
	2	479.4	95.7	4709	21	AAC69112	Human secreted pro
c	3	478.6	95.5	5065	21	AAC69111	Human secreted pro
c	4	478.6	95.5	5065	22	AAF33219	Human secreted pro
	5	477	95.2	1494	23	AAS68523	DNA encoding novel
	6	477	95.2	3192	23	AAS79105	DNA encoding novel
	7	477	95.2	3192	23	AAS81387	DNA encoding novel
	8	477	95.2	3432	23	AAS68186	DNA encoding novel
	9	477	95.2	3432	23	AAS70808	DNA encoding novel
	10	477	95.2	3432	23	AAS81679	DNA encoding novel
	11	477	95.2	3434	23	AAS87390	DNA encoding novel
	12	477	95.2	4137	23	AAS74444	DNA encoding novel
	13	477	95.2	4137	23	AAS78954	DNA encoding novel
	14	477	95.2	4137	23	AAS87865	DNA encoding novel
	15	477	95.2	4202	23	AAS73338	DNA encoding novel
	16	477	95.2	5367	23	AAS70936	DNA encoding novel
	17	477	95.2	7130	23	AAS78741	DNA encoding novel
	18	475.4	94.9	1785	23	AAS68656	DNA encoding novel
	19	475.4	94.9	1785	23	AAS79132	DNA encoding novel
	20	475.4	94.9	1919	23	AAS69603	DNA encoding novel
	21	475.4	94.9	1944	23	AAS77252	DNA encoding novel
	22	475.4	94.9	2025	23	AAS69612	DNA encoding novel
	23	475.4	94.9	2025	23	AAS73299	DNA encoding novel
	24	475.4	94.9	2025	23	AAS77455	DNA encoding novel
	25	475.4	94.9	2025	23	AAS78700	DNA encoding novel
	26	475.4	94.9	2025	23	AAS78928	DNA encoding novel
	27	475.4	94.9	2107	23	AAS74303	DNA encoding novel
	28	475.4	94.9	3558	23	AAS72863	DNA encoding novel
	29	475.4	94.9	3562	23	AAS79127	DNA encoding novel
	30	475.4	94.9	3657	23	AAS70241	DNA encoding novel
	31	475.4	94.9	3657	23	AAS73824	DNA encoding novel
	32	475.4	94.9	3657	23	AAS74283	DNA encoding novel
	33	475.4	94.9	3657	23	AAS74983	DNA encoding novel
	34	475.4	94.9	3657	23	AAS78725	DNA encoding novel
	35	475.4	94.9	3657	23	AAS78940	DNA encoding novel
	36	475.4	94.9	3657	23	AAS81661	DNA encoding novel
	37	475.4	94.9	3657	23	AAS84092	DNA encoding novel
	38	475.4	94.9	4251	23	AAS68182	DNA encoding novel
	39	475.4	94.9	4555	23	AAS84104	DNA encoding novel
	40	475.4	94.9	4797	23	AAS81681	DNA encoding novel
	41	475.4	94.9	5215	23	AAS70816	DNA encoding novel
	42	475.4	94.9	6194	23	AAS73339	DNA encoding novel
	43	475.4	94.9	13234	23	AAS82685	DNA encoding novel
c	44	475.4	94.9	1503841	24	ABT00010	Human neuregulin 1
c	45	475.4	94.9	1503841	24	ABT01503	Human neuregulin 1

ALIGNMENTS

RESULT 1
ABK47337
ID ABK47337 standard; DNA; 20300 BP.
XX
AC ABK47337;
XX
DT 18-JUN-2002 (first entry)
XX
DE Genomic nucleotide sequence encoding human RGS-4 protein.
XX
KW RGS-4; schizophrenia; human; regulator of G protein signalling 4;
KW neuroleptic; polymorphism; schizotypy; schizoid; chromosome 1q21-22;
KW gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace (4121,T)
FT /*tag= a
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4123,A)
FT /*tag= b
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4368,C)
FT /*tag= c
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4621,C)
FT /*tag= d
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace (4790,T)
FT /*tag= e

```
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (4816,T)
FT                   /*tag= f
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (4970,T)
FT                   /*tag= g
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace 5055,G)
FT                   /*tag= h
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (5295,A)
FT                   /*tag= i
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (5695,A)
FT                   /*tag= j
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (7375,T)
FT                   /*tag= k
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (7759,A)
FT                   /*tag= l
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (8596,A)
FT                   /*tag= m
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (9602..9610,CA)
FT                   /*tag= n
FT                   /note= "Deletion polymorphism"
FT   variation       replace (9892,A)
FT                   /*tag= o
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (9963,A)
FT                   /*tag= p
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (10132,A)
FT                   /*tag= q
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11056,C)
FT                   /*tag= r
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11091,T)
FT                   /*tag= s
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11106,A)
FT                   /*tag= t
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (11774,T)
FT                   /*tag= u
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (12143,A)
FT                   /*tag= v
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (12145,T)
FT                   /*tag= w
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (14367,G)
FT                   /*tag= x
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace (17027..17029,GA)
FT                   /*tag= y
FT                   /note= "Deletion polymorphism"
FT   variation       replace (17630,T)
FT                   /*tag= Z
FT                   /standard_name= "Single-nucleotide polymorphism"
XX
PN   WO200216653-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26622.
XX
PR   24-AUG-2000; 2000US-228021P.
XX
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Levitt PR, Mirnics K, Kodavali VC, Nimgaonkar VL;
XX
DR   WPI; 2002-292070/33.
XX
PT   Diagnosing, assessing susceptibility and treating schizophrenia,
PT   involves observing regulator of G-protein signalling 4, RGS4 levels in a
PT   subject -
XX
PS   Claim 1; Page 20-33; 112pp; English.
XX
CC   This invention relates to a novel method for diagnosing schizophrenia
CC   or determining susceptibility to schizophrenia in a human. The method
CC   comprises obtaining from a patient a DNA sample and detecting variations
CC   in the regulator of G-protein signalling 4 (RGS4) gene. Alternatively,
CC   the method involves measuring RGS4 mRNA or protein levels in a tissue
CC   sample from the patient and determining if there is a reduced level.
CC   The method of the invention is useful for diagnosing and determining
CC   susceptibility to schizophrenia. The invention also comprises a method
CC   that is useful for treating schizophrenia which includes a prophylactic
CC   treatment. The method of genotyping polymorphic variants in the RGS-4
CC   gene is applied to diagnosing pathologies of the schizophrenic spectrum,
CC   such as in particular schizotypy, schizoid individuals, etc. This
CC   method offers the possibility of diagnosing schizophrenia by a
CC   biological test and no longer exclusively by clinical evaluations.
CC   The present sequence represents the genomic DNA encoding the human
CC   regulator of G-protein signalling 4 (RGS4) protein used in the method of
CC   the invention. The gene for the RGS4 protein is located on human
CC   chromosome 1q21-22.
XX
SQ   Sequence 20300 BP; 6157 A; 4102 C; 3775 G; 6266 T; 0 other;

  Query Match             100.0%;  Score 501;  DB 24;  Length 20300;
  Best Local Similarity   100.0%;  Pred. No. 3e-133;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19800 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 19859

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19860 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 19919

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19920 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 19979

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19980 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 20039

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20040 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 20099

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20100 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 20159

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20160 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 20219

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20220 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 20279

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db      20280 CATGAGTGAACTCCCATTCAC 20300
```

```
RESULT 2
AAC69112
ID   AAC69112 standard; DNA; 4709 BP.
XX
AC   AAC69112;
XX
DT   31-JAN-2001  (first entry)
XX
DE   Human secreted protein gene 3 clone HNHCT15.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055371-A1.
XX
PD   21-SEP-2000.
XX
PF   16-MAR-2000; 2000WO-US06783.
XX
PR   18-MAR-1999;   99US-0125055.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ni J,  Ebner R,  Rosen CA,  Shi Y,  Birse C,  Florence K;
PI   Komatsoulis G,  Lafleur DW,  Moore PA,  Olsen HS,  Young PE;
XX
DR   WPI; 2000-594448/56.
DR   P-PSDB; AAB38012.
XX
PT   New nucleic acid molecules encoding 27 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives -
XX
PS   Claim 1; Page 385-386; 453pp; English.
XX
CC   The invention relate to the isolation of genes AAC69084-C69119 encoding
CC   27 human secreted proteins AAB37984-B38019. The genes can be used to
CC   generate fusion proteins by linking to the gene for the human
CC   immunoglobulin G Fc portion (AAC69075) for increasing the stability of
CC   the fusion protein as compared to the human protein only. The genes and
CC   proteins are useful for preventing, ameliorating or treating medical
CC   conditions, e.g. by protein or gene therapy. The genes are isolated
CC   from a range of human tissues disclosed in the specification. The
CC   nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
CC   wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
XX
SQ   Sequence 4709 BP; 1868 A; 990 C; 836 G; 939 T; 76 other;

  Query Match             95.7%;  Score 479.4;  DB 21;  Length 4709;
  Best Local Similarity   96.4%;  Pred. No. 2.7e-127;
  Matches  483;  Conservative    6;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db       2432 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACCCTTCAT 2491

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              |||||||||||||||| ||||||||||||||||:|||:||||||||||||  ||||

Db       2492 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACRTATYTCAAAATAATAAGAGCTAT 2551

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db       2552 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 2611

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2612 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 2671

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||||:|||||||||||||||||||||||||||||||||| ||||
Db       2672 AGTTCTGGCCAGGGCAATYAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 2731

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||:|||||||||||||||||||||||||||||:||
Db       2732 GGAAGTCAAATTGTCCCTGTTTGCAGAYGACATGATTGTATATCTAGAAAACCCCATYGT 2791

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2792 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 2851

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              ||||||:|||||||||||||||| ||||||||| ||||||||||||||||||||||||||
Db       2852 CAATGTRCAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 2911

Qy        481 CATGAGTGAACTCCCATTCAC 501
              |||||||||||||||||||||
Db       2912 CATGAGTGAACTCCCATTCAC 2932


RESULT 3
AAC69111/c
ID   AAC69111 standard; DNA; 5065 BP.
XX
AC   AAC69111;
XX
DT   31-JAN-2001  (first entry)
XX
DE   Human secreted protein gene 3 clone HPJCX13.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055371-A1.
XX
PD   21-SEP-2000.
XX
PF   16-MAR-2000; 2000WO-US06783.
XX
PR   18-MAR-1999;   99US-0125055.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ni J,  Ebner R,  Rosen CA,  Shi Y,  Birse C,  Florence K;
PI   Komatsoulis G,  Lafleur DW,  Moore PA,  Olsen HS,  Young PE;
XX
DR   WPI; 2000-594448/56.
DR   P-PSDB; AAB38011.
XX
PT   New nucleic acid molecules encoding 27 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives -
XX
PS   Claim 1; Page 383-384; 453pp; English.
XX
CC   The invention relate to the isolation of genes AAC69084-C69119 encoding
CC   27 human secreted proteins AAB37984-B38019. The genes can be used to
```

```
CC   generate fusion proteins by linking to the gene for the human
CC   immunoglobulin G Fc portion (AAC69075) for increasing the stability of
CC   the fusion protein as compared to the human protein only. The genes and
CC   proteins are useful for preventing, ameliorating or treating medical
CC   conditions, e.g. by protein or gene therapy. The genes are isolated
CC   from a range of human tissues disclosed in the specification. The
CC   nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
CC   wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
XX
SQ   Sequence 5065 BP; 980 A; 948 C; 1046 G; 2017 T; 74 other;

  Query Match             95.5%;  Score 478.6;  DB 21;  Length 5065;
  Best Local Similarity   96.6%;  Pred. No. 4.8e-127;
  Matches  484;  Conservative    4;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db       2312 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACCCTTCAT 2253

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              |||||||||||||||| ||||||||||||||||||||||||||||||||||  ||||
Db       2252 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 2193

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       2192 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 2133

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2132 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 2073

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||:||||||||||||||||||||||||||||||||| ||||
Db       2072 AGTTCTGGCCAGGGCAATYAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 2013

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||:||||||||||||||||||||||||||||:||
Db       2012 GGAAGTCAAATTGTCCCTGTTTGCAGAYGACATGATTGTATATCTAGAAAACCCCATYGT 1953

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1952 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 1893

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||| |||||||||||||| ||||||||| |||:||||||||||||||||||||||
Db       1892 CAATGTGCAAAAATCACAAGCATTCTTATACACCAAYAACAGACAAACAGAGAGCCAAAT 1833

Qy        481 CATGAGTGAACTCCCATTCAC 501
              |||||||||||||||||||||
Db       1832 CATGAGTGAACTCCCATTCAC 1812


RESULT 4
AAF33219/c
ID   AAF33219 standard; cDNA; 5065 BP.
XX
AC   AAF33219;
XX
DT   23-MAR-2001  (first entry)
XX
DE   Human secreted protein gene 7 SEQ ID NO:17.
XX
KW   Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
KW   dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW   immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
KW   ophthalmological; neuroprotectant; nootropic; anticonvulsant; vulnerary;
KW   antialzheimers; antiparkinsonian; antimicrobial; immune disorder;
KW   multiple sclerosis; systemic lupus erythematosus; HIV; infection;
KW   hyperproliferative disorder; cancer; Gaucher's disease; wound healing;
KW   cardiovascular disease; Scimitar syndrome; Chaga's cardiomyopathy;
KW   coronary arteriosclerosis; angiogenic disorder; diabetic retinopathy;
KW   corneal graft neovascularisation; neurological disorder; regeneration;
KW   Huntington's chorea; Alzheimer's disease; Parkinson's disease;
KW   infectious disease; chemotaxis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200076530-A1.
XX
PD   21-DEC-2000.
XX
PF   01-JUN-2000; 2000WO-US14933.
XX
PR   11-JUN-1999;   99US-0138572.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis GA;
XX
DR   WPI; 2001-071147/08.
DR   P-PSDB; AAB64888.
XX
PT   Nucleic acids encoding 49 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT   disease and diabetic retinopathy -
XX
PS   Claim 1; Page 453-455; 554pp; English.
XX
CC   The polynucleotide sequences given in AAF33213 to AAF33261 encode the
CC   human secreted proteins given in AAB64882 to AAB64930. AAB64931 to
CC   AAB64991 represent human secreted polypeptide sequences and proteins
CC   homologous to them, which are given in the exemplification of the present
CC   invention. Human secreted proteins have activities based on the tissues
CC   and cells the genes are expressed in. Examples of activities include:
CC   immunomodulatory; antisclerotic; dermatological; immunosuppressive;
CC   antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;
CC   vascular; antimicrobial; anti-angiogenic; ophthalmological;
CC   neuroprotectant; anticonvulsant; nootropic; antialzheimers;
CC   antiparkinsonian; and vulnerary. The polynucleotides and polypeptides can
CC   be used in the prevention, diagnosis and treatment of diseases associated
CC   with inappropriate polypeptide expression. Disorders that may be
CC   prevented, diagnosed and/or treated by the above methods include immune
CC   disorders (e.g. multiple sclerosis, systemic lupus erythematosus and
CC   human immuno-deficiency virus (HIV) infections), hyperproliferative
CC   disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
CC   (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
CC   arteriosclerosis), angiogenic disorders (e.g. corneal graft
CC   neovascularisation and diabetic retinopathy), neurological disorders
CC   (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
CC   infectious diseases and/or for promoting wound healing, regeneration and
CC   /or chemotaxis. AAF33204 to AAF33212 and AAB64881 represent sequences
CC   used in the exemplification of the present invention.
XX
SQ   Sequence 5065 BP; 980 A; 948 C; 1046 G; 2017 T; 74 other;

  Query Match             95.5%;  Score 478.6;  DB 22;  Length 5065;
  Best Local Similarity   96.6%;  Pred. No. 4.8e-127;
  Matches  484;  Conservative    4;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db       2312 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACCCTTCAT 2253

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              |||||||||||||||| ||||||||||||||||||||||||||||||||||  ||||
Db       2252 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 2193
```

```
Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db       2192 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 2133

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2132 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 2073

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || ||||||||||||||:|||||||||||||||||||||||||||||||| ||||
Db       2072 AGTTCTGGCCAGGGCAATYAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 2013

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||:|||||||||||||||||||||||||||||:||
Db       2012 GGAAGTCAAATTGTCCCTGTTTGCAGAYGACATGATTGTATATCTAGAAAACCCCATYGT 1953

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1952 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 1893

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||| |||||||||||||| ||||||||| |||:|||||||||||||||||||||||
Db       1892 CAATGTGCAAAAATCACAAGCATTCTTATACACCAAYAACAGACAAACAGAGAGCCAAAT 1833

Qy        481 CATGAGTGAACTCCCATTCAC 501
              |||||||||||||||||||||
Db       1832 CATGAGTGAACTCCCATTCAC 1812
```

RESULT 5
AAS68523
ID AAS68523 standard; cDNA; 1494 BP.
XX
AC AAS68523;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #4327.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG04336.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 4327; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1494 BP; 569 A; 318 C; 277 G; 330 T; 0 other;

Query Match 95.2%; Score 477; DB 23; Length 1494;
Best Local Similarity 97.0%; Pred. No. 8.6e-127;
Matches 486; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              |||||||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db        495 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACCCTTCAT 554

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              |||||||||||||||| ||||||||| |||||||||||||||||||||||||  ||||
Db        555 GCTAAAAACTCTCAATAAATTAGGTACTGATGGGACGTATCTCAAAATAATAAGAGCTAT 614

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db        615 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 674

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        675 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 734

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||||||||||||||||||||||||| |||||||||||||| ||||
Db        735 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAAGGTATTCAATTAGGAAAAGA 794

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        795 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGT 854

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        855 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 914

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||| ||||||||| |||||||||||||||||||||||||||||
Db        915 CAATGTACAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 974

Qy        481 CATGAGTGAACTCCCATTCAC 501
              |||||||||||||||||||||
Db        975 CATGAGTGAACTCCCATTCAC 995
```

RESULT 6
AAS79105
ID AAS79105 standard; cDNA; 3192 BP.
XX
AC AAS79105;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #14909.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX

```
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG14918.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 14909; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3192 BP; 1290 A; 705 C; 562 G; 635 T; 0 other;

  Query Match             95.2%;  Score 477;  DB 23;  Length 3192;
  Best Local Similarity   97.0%;  Pred. No. 1.2e-126;
  Matches  486;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||| |||||||||||||| |||||| ||||||
Db       1353 CCACATGATTATCTCAATAGATGCAGAAAAAGCATTTGACAAAATTCAACAACCCTTCAT 1412

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              |||||||||||||||| ||||||||||||||||||||||||||||||||||  ||||
Db       1413 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 1472

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1473 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 1532

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1533 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 1592

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       1593 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 1652

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||| ||
Db       1653 GGAAGTCAAATTGTCCCTGTTTGCAGACGACATGATTGTATATCTAGAAAACCCCATTGT 1712

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1713 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 1772

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||| ||||||||| ||| ||||||||||||||||||||||||
Db       1773 CAATGTACAAAAATCACAAGCATTCTTATACACCAACAACAGACAAACAGAGAGCCAAAT 1832

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db       1833 CATGAGTGAACTCCCATTCAC 1853


RESULT 7
AAS81387
ID   AAS81387 standard; cDNA; 3192 BP.
XX
AC   AAS81387;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #17191.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG17200.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 17191; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
```

```
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3192 BP; 1290 A; 705 C; 562 G; 635 T; 0 other;

  Query Match             95.2%;  Score 477;  DB 23;  Length 3192;
  Best Local Similarity   97.0%;  Pred. No. 1.2e-126;
  Matches  486;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||| ||||||||||||||| |||||| ||||||
Db       1353 CCACATGATTATCTCAATAGATGCAGAAAAAGCATTTGACAAAATTCAACAACCCTTCAT 1412

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              ||||||||||||||| |||||||||||||||||||||||||||||||  ||||
Db       1413 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 1472

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1473 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 1532

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1533 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 1592

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       1593 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 1652

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||| ||
Db       1653 GGAAGTCAAATTGTCCCTGTTTGCAGACGACATGATTGTATATCTAGAAAACCCCATTGT 1712

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1713 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 1772

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||||| ||||||||| ||| ||||||||||||||||||||||
Db       1773 CAATGTACAAAAATCACAAGCATTCTTATACACCAACAACAGACAAACAGAGAGCCAAAT 1832

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db       1833 CATGAGTGAACTCCCATTCAC 1853


RESULT 8
AAS68186
ID   AAS68186 standard; cDNA; 3432 BP.
XX
AC   AAS68186;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #3990.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG03999.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 3990; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3432 BP; 1349 A; 769 C; 624 G; 690 T; 0 other;

  Query Match             95.2%;  Score 477;  DB 23;  Length 3432;
  Best Local Similarity   97.0%;  Pred. No. 1.2e-126;
  Matches  486;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||| |||||||||||||||| |||||| ||||||
Db       1353 CCACATGATTATCTCAATAGATGCAGAAAAAGCATTTGACAAAATTCAACAACCCTTCAT 1412

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              ||||||||||||||| |||||||||||||||||||||||||||||||||||  ||||
Db       1413 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 1472

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1473 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 1532

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1533 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 1592

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       1593 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 1652

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||| ||
Db       1653 GGAAGTCAAATTGTCCCTGTTTGCAGACGACATGATTGTATATCTAGAAAACCCCATTGT 1712

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1713 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 1772

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||||| ||||||||| ||| ||||||||||||||||||||||
Db       1773 CAATGTACAAAAATCACAAGCATTCTTATACACCAACAACAGACAAACAGAGAGCCAAAT 1832

Qy        481 CATGAGTGAACTCCCATTCAC 501
```

```
                    ||||||||||||||||||||
Db       1833 CATGAGTGAACTCCCATTCAC 1853


RESULT 9
AAS70808
ID   AAS70808 standard; cDNA; 3432 BP.
XX
AC   AAS70808;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #6612.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG06621.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 6612; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3432 BP; 1349 A; 769 C; 624 G; 690 T; 0 other;

  Query Match             95.2%;  Score 477;  DB 23;  Length 3432;
  Best Local Similarity   97.0%;  Pred. No. 1.2e-126;
  Matches  486;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||| ||||||||||||||| |||||| ||||||
Db       1353 CCACATGATTATCTCAATAGATGCAGAAAAAGCATTTGACAAAATTCAACAACCCTTCAT 1412

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||  ||||
Db       1413 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 1472

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1473 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 1532

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1533 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 1592

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       1593 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 1652

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||| ||
Db       1653 GGAAGTCAAATTGTCCCTGTTTGCAGACGACATGATTGTATATCTAGAAAACCCCATTGT 1712

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1713 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 1772

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||| ||||||||| ||| ||||||||||||||||||||||||
Db       1773 CAATGTACAAAAATCACAAGCATTCTTATACACCAACAACAGACAAACAGAGAGCCAAAT 1832

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db       1833 CATGAGTGAACTCCCATTCAC 1853


RESULT 10
AAS81679
ID   AAS81679 standard; cDNA; 3432 BP.
XX
AC   AAS81679;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #17483.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG17492.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 17483; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
```

```
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3432 BP; 1349 A; 769 C; 624 G; 690 T; 0 other;

  Query Match             95.2%;  Score 477;  DB 23;  Length 3432;
  Best Local Similarity   97.0%;  Pred. No. 1.2e-126;
  Matches  486;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              |||||||||||||||||||||||||||||| ||||||||||||||| |||||| ||||||
Db       1353 CCACATGATTATCTCAATAGATGCAGAAAAAGCATTTGACAAAATTCAACAACCCTTCAT 1412

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||  ||||
Db       1413 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 1472

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1473 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 1532

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1533 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 1592

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       1593 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 1652

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||| ||
Db       1653 GGAAGTCAAATTGTCCCTGTTTGCAGACGACATGATTGTATATCTAGAAAACCCCATTGT 1712

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1713 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 1772

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||| |||||||| ||| ||||||||||||||||||||||||||
Db       1773 CAATGTACAAAAATCACAAGCATTCTTATACACCAACAACAGACAAACAGAGAGCCAAAT 1832

Qy        481 CATGAGTGAACTCCCATTCAC 501
              |||||||||||||||||||||
Db       1833 CATGAGTGAACTCCCATTCAC 1853


RESULT 11
AAS87390
ID   AAS87390 standard; cDNA; 3434 BP.
XX
AC   AAS87390;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #23194.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG23203.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 23194; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3434 BP; 1349 A; 771 C; 624 G; 690 T; 0 other;

  Query Match             95.2%;  Score 477;  DB 23;  Length 3434;
  Best Local Similarity   97.0%;  Pred. No. 1.2e-126;
  Matches  486;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              |||||||||||||||||||||||||||||| ||||||||||||||| |||||| ||||||
Db       1355 CCACATGATTATCTCAATAGATGCAGAAAAAGCATTTGACAAAATTCAACAACCCTTCAT 1414

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||  ||||
Db       1415 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 1474

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1475 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 1534

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1535 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 1594
```

```
Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       1595 AGTTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 1654

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||| ||
Db       1655 GGAAGTCAAATTGTCCCTGTTTGCAGACGACATGATTGTATATCTAGAAAACCCCATTGT 1714

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1715 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 1774

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||||| ||||||||| ||| ||||||||||||||||||||||
Db       1775 CAATGTACAAAAATCACAAGCATTCTTATACACCAACAACAGACAAACAGAGAGCCAAAT 1834

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db       1835 CATGAGTGAACTCCCATTCAC 1855
```

RESULT 12
AAS74444
ID AAS74444 standard; cDNA; 4137 BP.
XX
AC AAS74444;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #10248.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG10257.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 10248; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4137 BP; 1397 A; 980 C; 935 G; 825 T; 0 other;

Query Match 95.2%; Score 477; DB 23; Length 4137;
Best Local Similarity 97.0%; Pred. No. 1.3e-126;
Matches 486; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              |||||||||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db       2265 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACCCTTCAT 2324

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||  ||||
Db       2325 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 2384

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       2385 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 2444

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2445 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 2504

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||| ||||||||||||||||||||||||||||||||||||| ||||
Db       2505 AGTTCTGGCCAGGGCAATCAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 2564

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db       2565 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGT 2624

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       2625 CTCAGCCCAAAATCCCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 2684

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||
Db       2685 CAATGTACAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 2744

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db       2745 CATGAGTGAACTCCCATTCAC 2765
```

RESULT 13
AAS78954
ID AAS78954 standard; cDNA; 4137 BP.
XX
AC AAS78954;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #14758.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.

```
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG14767.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 14758; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4137 BP; 1398 A; 982 C; 934 G; 823 T; 0 other;

  Query Match             95.2%;  Score 477;  DB 23;  Length 4137;
  Best Local Similarity   97.0%;  Pred. No. 1.3e-126;
  Matches  486;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db       2265 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACCCTTCAT 2324

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              |||||||||||||||| |||||||||||||||||||||||||||||||||  ||||
Db       2325 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 2384

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db       2385 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 2444

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2445 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 2504

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||| |||||||||||||||||||||||||||||||| ||||
Db       2505 AGTTCTGGCCAGGGCAATCAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 2564

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db       2565 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGT 2624

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       2625 CTCAGCCCAAAATCCCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 2684

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||
Db       2685 CAATGTACAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 2744

Qy        481 CATGAGTGAACTCCCATTCAC 501
              ||||||||||||||||||||
Db       2745 CATGAGTGAACTCCCATTCAC 2765


RESULT 14
AAS87865
ID   AAS87865 standard; cDNA; 4137 BP.
XX
AC   AAS87865;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #23669.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG23678.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 23669; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4137 BP; 1398 A; 982 C; 934 G; 823 T; 0 other;

  Query Match             95.2%;  Score 477;  DB 23;  Length 4137;
  Best Local Similarity   97.0%;  Pred. No. 1.3e-126;
  Matches  486;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;
```

```
Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db       2265 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACCCTTCAT 2324

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||  ||||
Db       2325 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 2384

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db       2385 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 2444

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2445 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 2504

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||| |||||||||||||||||||||||||||||||||||| ||||
Db       2505 AGTTCTGGCCAGGGCAATCAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 2564

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db       2565 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGT 2624

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db       2625 CTCAGCCCAAAATCCCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 2684

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||
Db       2685 CAATGTACAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 2744

Qy        481 CATGAGTGAACTCCCATTCAC 501
              |||||||||||||||||||||
Db       2745 CATGAGTGAACTCCCATTCAC 2765
```

RESULT 15
AAS73338
ID AAS73338 standard; cDNA; 4202 BP.
XX
AC AAS73338;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #9142.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG09151.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 9142; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4202 BP; 1482 A; 951 C; 928 G; 841 T; 0 other;

Query Match 95.2%; Score 477; DB 23; Length 4202;
Best Local Similarity 97.0%; Pred. No. 1.3e-126;
Matches 486; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
Qy          1 CCACATGATTATCTCAATAGATGCAGAAAAGGCATTTGACAAAATTTAACAACTCTTCAT 60
              ||||||||||||||||||||||||||||||||| |||||||||||| |||||| ||||||
Db       2265 CCACATGATTATCTCAATAGATGCAGAAAAGGCCTTTGACAAAATTCAACAACCCTTCAT 2324

Qy         61 GCTAAAAACTCTCAATCAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGCACTAT 120
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||  ||||
Db       2325 GCTAAAAACTCTCAATAAATTAGGTATTGATGGGACGTATCTCAAAATAATAAGAGCTAT 2384

Qy        121 CTATGACAAACTCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 180
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db       2385 CTATGACAAACCCACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGCATTCCCTTT 2444

Qy        181 GAAAACGGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 240
              |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2445 GAAAACTGGCACAAGACAGGGATGCCCTCTCTCACCACTCCTATTCAACATAGTGTTGGA 2504

Qy        241 AGCTCTGGCCAGGGCAATTAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAGAAGA 300
              || |||||||||||||| |||||||||||||||||||||||||||||||||||| ||||
Db       2505 AGTTCTGGCCAGGGCAATCAGGCAGGAGAAGGAAATAAAGGGTATTCAATTAGGAAAAGA 2564

Qy        301 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATCGT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db       2565 GGAAGTCAAATTGTCCCTGTTTGCAGATGACATGATTGTATATCTAGAAAACCCCATTGT 2624

Qy        361 CTCAGCCCAAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 420
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db       2625 CTCAGCCCAAAATCCCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTCAGGATACAAAAT 2684

Qy        421 CAATGTACAAAAATCACAAGCACTCTTATACATCAATAACAGACAAACAGAGAGCCAAAT 480
              |||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||
Db       2685 CAATGTACAAAAATCACAAGCATTCTTATACACCAATAACAGACAAACAGAGAGCCAAAT 2744

Qy        481 CATGAGTGAACTCCCATTCAC 501
              |||||||||||||||||||||
Db       2745 CATGAGTGAACTCCCATTCAC 2765
```

Search completed: November 7, 2003, 06:15:47
Job time : 158.728 secs